6-2017

THE SANTA ANA SPECKLED DACE (RHINICHTHYS OSCULUS): PHYLOGEOGRAPHY AND MOLECULAR EVOLUTION OF THE MITOCHONDRIAL DNA CONTROL REGION

James Jay VanMeter
California State University San Bernardino, piaandjayv@sbcglobal.net

THE SANTA ANA SPECKLED DACE (*RHINICHTHYS OSCULUS*): PHYLOGEOGRAPHY AND MOLECULAR EVOLUTION OF THE MITOCHONDRIAL DNA CONTROL REGION

_____

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

_____

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

in

Biology

_____

by

James Jay VanMeter

June 2017

THE SANTA ANA SPECKLED DACE (*RHINICHTHYS OSCULUS*): PHYLOGEOGRAPHY AND MOLECULAR EVOLUTION OF THE MITOCHONDRIAL DNA CONTROL REGION

A Thesis

Presented to the

Faculty of

California State University,

San Bernardino

---

by

James Jay VanMeter

June 2017

Approved by:

Anthony Metcalf, Committee Chair, Biology

James Ferrari, Committee Member

David Polcyn, Committee Member

ABSTRACT

The purpose of this genetic study of the Santa Ana Speckled Dace *Rhinichthys osculus* was three-fold. The first goal was to characterize the molecular structure of the mtDNA control region of *R. osculus*. An 1143 base-pair region of the mitochondrial DNA genome, which included the complete control region was sequenced for all individuals. Analysis of the sequence data revealed that the molecular structure of the speckled dace control region was similar to the molecular structure described for other vertebrate taxa. The speckled dace control region contains three major domains, which vary in base frequency as well as in the frequency of nucleotide polymorphisms. Domain II was observed to be the most conserved, and Domain I was the most variable domain of the control region, in agreement with studies of other vertebrate control regions. .

The second goal of this study was to ascertain the phylogeny of *R. osculus* in Southern California in relation to other speckled dace in California. Seventy-four specimens of *R. osculus* were collected from five different watersheds located in three geographic regions of California: Southern California, the Central California Coast, and the Eastern Desert of the Owens River valley. Phylogenetic analysis of sequence data revealed that the Santa Ana Speckled Dace is a genetically distinct population from *R. osculus* inhabiting the Central Coast or Eastern Desert regions, which both differ from the Santa Ana Speckled Dace by a genetic distance of more than 7 percent. The Santa Ana

Speckled Dace inhabiting the watersheds of Southern California form a reciprocally monophyletic clade with respect to the Central Coast dace and the Eastern Desert dace, which are sister clades to one another.

The third goal of this study was to describe the population genetics of *R. osculus* in Southern California. Population genetic analysis demonstrated that a high degree of geographic population structure exists for the Speckled Dace in California, with 96% of molecular variance attributable to regional differences through isolation by distance. A high degree of population structure also exists among populations within the Southern California region as well. It was found that 45% of molecular variance in the Santa Ana Speckled Dace is attributable to differences among tributaries. This study finds that the distribution of speckled dace in Southern California best fits a model of population structure by individual tributary, with episodes of localized population bottlenecks followed by sudden population expansion, most likely linked to climatic variation.

It is proposed the Santa Ana Speckled Dace constitutes an Evolutionarily Significant Unit (ESU) that qualifies it as a separate subspecies on the basis that it is geographically separated from other populations, genetically distinct from other dace populations due to restricted gene flow, and possesses unique phenotypic characteristics. This information suggests conservation and management strategies for the speckled dace populations which remain in existence in Southern California.

ACKNOWLEDGMENTS

I would like to thank the many people who provided guidance, advice, and support in the time this study was completed at CSUSB. First, I would like to express my thanks to my first lab mates Tom Benson, George Walker, and Keri Maher. Tom provided help to Dr. Anthony Metcalf to collect the specimens making up the bulk of this study, which are also being used in other studies of the speckled dace as well. George did the initial sequencing of dace specimens, and thanks to his work an improved technique to obtain DNA sequence data was discovered. George was also the person who instructed my wife Pia and me how to perform sequencing reactions and resolve and assemble contigs into an amplicon, and I appreciate the late nights he stayed in the lab to teach us those techniques. I thank Keri for her tips about good technique, and her good humor. I would especially like to thank my advisor Dr. Anthony Metcalf for his amazing expertise and knowledge, which he was always willing to share freely. I am especially grateful to Dr. Metcalf because of his unlimited patience regarding the many years it took to complete this project. I would like to thank Dr. James Ferrari and Dr. David Polcyn, who formed my thesis committee and were always supportive of me no matter the circumstances. I am indebted to you both because of your kind patience. I thank my lab mates Stacy Nerkowski and Lianne Greaver for their good humor and the great decorations they have added to the lab. I thank my mother Pat VanMeter for her unqualified love and support through my entire life and for her support for my education from grade school to graduate

school. Finally I would like to thank my wife, lab mate, and the love of my life Pia VanMeter, who was also a graduate biology student working on the genetics of the Santa Ana Speckled Dace at the same time I worked on this project. Without her inspiration, I would never have applied to graduate school and had the opportunity to work on this project. Thank you Pia, you made every day in the lab fun (as you make every day in my life fun!), and I always looked forward to the time we spent together all those summers, late weeknights after work, and weekends doing the work for this thesis.

My research was supported by grants and assistance from the Associated Students Incorporated at California State University San Bernardino, the California Department of Fish and Wildlife, and the United States Forest Service.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGEMENTS v

CHAPTER ONE: INTRODUCTION

Overview 1

Phylogeographic Studies 4

Methods of Phylogeny Reconstruction 5

Phylogeography of *Rhinichthys osculus* 8

MtDNA Phylogeography of Freshwater Vertebrates 10

Phylogeography of Southern California Vertebrates 16

Mitochondrial DNA as a Molecular Marker 19

Mitochondrial DNA: Structure and Function in Vertebrates 20

Structure and Evolution of the Mitochondrial DNA Control Region 21

Population Genetics 29

Evolutionary Forces and Population Differentiation 31

Genetic Structure of Populations and Gene Flow 33

Proposed Study 34

CHAPTER TWO: MATERIALS AND METHODS

Data Collection 37

Molecular Methods 38

Sequence Analysis 41

Phylogeography 42

Population Genetics 44
Summary 46
CHAPTER THREE: RESULTS
Sequence Analysis 48
Structure of the mtDNA Control Region 49
Phylogeography and Population Structure 52
Population Genetics 54
CHAPTER FOUR: DISCUSSION 60
Structure of the mtDNA Control Region of Speckled Dace 61
Phylogeography and Population Structure 65
Phylogeographic Analysis 70
Santa Ana Speckled Dace 74
Conservation Implications 75
APPENDIX A: FIGURES 77
APPENDIX B: TABLES 96
APPENDIX C: INPUT FILES 107
REFERENCES 160

CHAPTER ONE

INTRODUCTION

Overview

Vertebrate animal populations are useful models of study to ascertain the occurrence and degree of genetic differentiation across landscapes, and to estimate the rates of evolution of the genes under study. The distribution of genetic variation within and among populations illustrates the interaction between the geographic variables of climate and climatic change, and the effect physical barriers have on gene flow. Southern California affords a unique opportunity to examine this interplay, as it is part of one of the most diverse faunal assemblages extant, the "California Floristic Province." (Hickman, 1993; Myers et al, 1999).

The complex interaction between the geography of the region and climatic events has influenced the evolution of many Southern California species (Rodriguez-Robles et al, 1999, 2001; Maldonado et al, 2001; Metcalf et al, 2001; Sgariglia and Burns, 2003; reviewed in Calsbeek et al, 2003; Forister et al, 2004; Phillipsen, 2005; Chatzimanolis and Caterino, 2007; Phillipsen and Metcalf, 2009). In particular, stream-dwelling freshwater fish are ideal subjects of vertebrate population genetic studies, owing to their restricted vagility and relative isolation within small areas of suitable habitat (Meffe and Vrijenhoek 1988; Avise 2000). The precarious existence such freshwater organisms lead

closely ties their evolutionary history at the levels of both population and species to the geologic history of the landscape they inhabit (Bernatchez and Wilson 1998).

Among the most successful freshwater fish are the family Cyprinidae (minnows), members of which are found in virtually every major freshwater drainage in western North America (Lee et al, 1980). Study of the morphology and paleontology of cyprinid fish has been employed in the past to reconstruct the biogeography of portions of western North America (Hubbs and Miller 1948; Minckley et al, 1986). A member of the Cyprinidae, the species *Rhinichthys osculus* is among the most widely distributed fish species in the western United States and the only fish genus native to all major western drainages from Canada south to Sonora, Mexico (Lee et al, 1980; Hocutt, 1986; Moyle 2002). Within Southern California *Rhinichthys osculus* is known locally as the subspecies Santa Ana Speckled Dace, *Rhinichthys osculus carringtonii* (Jordan and Evermann, 1896; Culver and Hubbs, 1917; McGinnis, 2006). The Santa Ana speckled dace is a small cyprinid, usually less than 80 mm total length (Figure 1, appendix A). Speckled dace typically inhabit streams containing shallow cobble and gravel riffles, with overhanging riparian plants providing cover for the fish from predators (Moyle, 2002). The Santa Ana speckled dace was once widely distributed throughout the upland sections of the Santa Ana, San Gabriel, and Los Angeles river systems, but presently inhabits small isolated streams and creeks in the Santa Ana and San Gabriel river systems. The habitat presently

occupied by the speckled dace elsewhere throughout California is similarly discontinuous (Figure 2, appendix A). It was reported the Santa Ana Speckled Dace has been completely extirpated from the Los Angeles River drainage system (SAWPA, 2004), although specimens of *R. osculus* were collected from Big Tujunga Creek for a study as recently as 2008 (Feeney and Swift, 2008). The local range occupied by *Rhinichthys osculus* is restricted due to habitat fragmentation attributed to human activity (Feeney and Swift, 2008) and to the aftermath of the effects of fire and flood (SAWPA 2004; Santa Ana Speckled Dace Recovery Project 2005). The Santa Ana Speckled Dace was listed as a species of special concern in 1995 by the California Department of Fish and Game and as a sensitive species by the United States Forest Service in 1998. However, listing as a federally endangered species was denied due to a lack of formal peer reviewed description (Moyle, 2002).

This study proposes to characterize the distribution of genetic variation in *Rhinichthys osculus* in Southern California at the subpopulation, population, and watershed levels in order to reveal the true phylogeny of *R. osculus* within this geographic area. The method employed in this proposal will involve examination of mitochondrial DNA (mtDNA) sequence data. As described below, mtDNA is the most appropriate molecule to determine the phylogeography of *Rhinichthys osculus.* Utilizing mtDNA will also afford the opportunity to study the evolution of a molecular marker at the population level. This would lead to a better understanding of how local geography, climate, and migration have affected the

evolution of the species within the region, as well as contribute to the overall body of knowledge regarding the phylogeography of Southern California species in general, and freshwater vertebrates in particular.

Phylogeographic Studies

The basic aim of phylogeography is to infer the phylogeny and historical evolution of a particular species from its observed geographic distribution of genetic variation, in concordance with known historical geological and ecological events (Avise et al, 1987; Avise 1998; Avise 2000). Phylogenetic studies often utilize molecular genetic data taken from individuals forming a population in the form of mitochondrial DNA due to its advantageous properties in most cases (in contrast to single copy nuclear DNA, or scnDNA) of clonal (Hayashi et al, 1985) and haploid uniparental inheritance (Dawid et al, 1972) to reconstruct maternal lineages, or gene trees (Moritz et al, 1987; Harrison, 1989). Due to the lack of recombination of mtDNA in vertebrate species, genetic variation is introduced through mutation alone. An advantage of uniparental haploid inheritance of mtDNA is that effective population size (N_e) as estimated using mtDNA data is one-fourth N_e estimated using scnDNA data (Birky et al, 1983). Small effective population size has the advantage of shorter coalescent time, or time before a population drifts toward one predominant mtDNA lineage (Moore, 1995). Thus, the degree of genetic drift and geographic variation among populations is more likely to reflect their divergence times, so long as saturation is not significant

(Chenoweth et al, 1998; Burton and Lee, 1994; Palumbi and Baker, 1994). This results in an increased likelihood that a gene tree based on mtDNA data will also represent the true phylogeny of the organism.

Phylogenetic approaches employ the methods of systematics to draw phylogenetic inferences for the taxa under study (Hennig, 1966; Mayden and Wiley, 1992). Using the technique of polymerase chain reaction (PCR) (White et al, 1989), select portions of the mtDNA genome may be amplified *in vitro* and the resulting amplicon directly sequenced (Scharf et al, 1986; Sanger et al, 1977). Analysis of the complete sequence of a particular gene in different populations of higher taxa may also elucidate the molecular evolution of the gene itself. This information, in turn, may then feed back into the evolutionary analysis and phylogeography of the species in question. Thus, molecular analysis of nucleotide sequences may generate both character-state and distance-based genetic markers to answer questions concerning a species' population structure, range, gene flow, and molecular evolution.

Methods of Phylogeny Reconstruction

The method of phylogeny reconstruction by genetic distance utilizes genetic distance data plotted in matrix form to generate gene trees, which are constructed from pair wise distance values by one of two general methods, neighbor-joining (N-J) or minimum evolution (ME). The N-J approach connects nearest neighbors within the matrix together until all haplotypes are included in

the resulting tree, while ME seeks to produce the tree with the minimum overall genetic distance.

The method of maximum parsimony (MP) is a character-state based approach rather than a distance-based approach as described above. Maximum parsimony seeks to construct gene trees with the requirement that the tree be based on the minimum number of evolutionary steps or changes necessary to explain the differences in two related DNA sequences. The total number of steps required to traverse the entire tree is determined, and the tree(s) with the overall minimum total number of evolutionary steps is taken to be the most parsimonious.

A disadvantage to both the ME and MP methods is that they do not take into account that DNA substitutions may not all occur at the same rate at all positions, or that nucleotide bias may exist. The simplest model of molecular evolution is one in which all DNA nucleotide base frequencies are equal, and all substitutions occur at the same rate, which is known as the Jukes-Cantor model (Jukes and Cantor, 1969). More complex models of molecular evolution take into account differences in nucleotide base frequencies and differential rates of nucleotide substitution, for example that the rate for transitions (purine to purine, or pyrimidine to pyrimidine) may differ from the rate for transversions (purine to pyrimidine, or the reverse) (Tamura, 1992). Evolutionary models that best fit observed data are determined using software programs MEGA 6.0 (Tamura et al. 2013) and Modeltest 3.06 (Posada and Crandall, 1998).

The method of maximum likelihood (ML) gene tree construction selects the gene tree that best fits the genetic data given an evolutionary model, which is chosen based on analysis of the patterns of substitutions observed in the DNA sequence data described above. ML methods assume that DNA changes are more likely for long branches (high divergence) between taxa than for short branches (low divergence), and so branch length estimation plays an important part of this method. One of the advantages of the ML approach is that it is the least susceptible method of gene tree construction to sampling error (Swofford et al, 1996).

The Bayesian method of inferring phylogeny employs a Markov chain Monte Carlo (MCMC) likelihood algorithm, and a set of prior probabilities based on a model of evolution which fits the observed data, to estimate a posterior probability that a particular phylogenetic tree is the most likely tree for the given data set. Bayesian inference of phylogeny has become the major method to reconstruct phylogenies due to its computational efficiency, ability to assess the robustness of uncertainty, and ability to incorporate different models of evolution in the analysis (Ronquist et al., 2012).

Character state-based methods are most often employed to generate phylogenetic trees or cladograms of maximum parsimony or maximum likelihood from mtDNA control region sequence data. Several recent examples of phylogeny reconstruction employing fish mtDNA control region sequence data include inference of the phylogeography of cyprinid fish in Southeast Asia

(McConnell, 2004), revealing the extant major lineages of east Asian ninespine stickleback (Takahashi and Goto, 2001), determination of the genetic population structure of blacktip sharks in the northwest Atlantic, Gulf of Mexico, and Caribbean Sea (Keeney et al., 2005), phylogenetic analysis of grayling (genus: *Thymallus*) on the Eurasian (Froufe et al., 2005) and European (Weiss et al., 2013) continents, inferring the colonization pattern of European cyprinids from glacial refugia (Costedoat et al., 2006), and detecting population expansion of a Caribbean reef fish *H. bivittatus* (Haney et al., 2010).

Phylogeography of *Rhinichthys osculus*

Previous descriptions of *Rhinichthys* have focused primarily on morphology to characterize the genus (Jordan and Evermann, 1896; Hubbs and Miller, 1948; Cornelius, 1969; Hubbs et al, 1974; Woodman 1992; Moyle 2002; Fraser et al, 2005). Although extensive study of the molecular genetics of Cyprinids in Europe has been reported, few North American studies at the molecular level have been described in the literature. Speckled dace are known to exhibit genetic differentiation at the subpopulation, population, and regional levels (Pfrender et al 2004; Oakey et al, 2004). Early morphological descriptions of *Rhinichthys osculus* characterized it as being comprised of 12 species (Jordan and Evermann, 1896), while more recently this number has been reduced to one species with multiple localized subspecies that result from a complex interaction of geotectonic and climatic events (Hubbs et al 1974). Cornelius (1969)

concluded that the local subspecies *R. osculus carringtonii*, the Santa Ana Speckled Dace, was more closely related to *R. osculus yarrowi* from the Virgin River in Arizona, a Middle Colorado River tributary, than to populations of *R. osculus carringtonii* from North-Coastal California based on morphometric and meristic analysis. Oakey et al. (2004) arrived at a similar conclusion based on molecular data utilizing RFLP analysis of mtDNA from *R. osculus.* This similarity may be explained by hypothesized intermittent connections between the Colorado River system and the Los Angeles Basin via the Mohave River during the late Pliocene and Pleistocene pluvial periods, prior to the Transverse ranges attaining their present altitude (Miller, 1946; Cornelius, 1969; Smith and Dowling, 2008), or from the Colorado River delta system via migration and subsequent headwater capture into the Pliocene Los Angeles Basin when it occupied a location more southerly than at present due to relative motion of the San Andreas Fault system (Spencer et al, 2008).

Vertebrate animal taxa inhabiting the region of Southern California have been subject to a complex series of events shaping their evolutionary history. A brief chronology of events within the recent geologic past includes establishment of the San Andreas and associated fault systems 20 – 10 million years before present(or m.y.b.p.), secondary uplift of the Sierra Nevada Mountains 10 – 5 m.y.b.p., and the formation of the Peninsular, Transverse, and Coast Mountain ranges 5 – 3 m.y.b.p. (Hickman, 1993). More recently, following the end of the final pluvial period of the Pleistocene approximately 11,000 y.b.p., the climate of

the Southern California region has been subject to increased aridification (Harden, 2004; Thompson et al, 1993). The effect of such geologic and climatic forces acting either alone or in concert may potentially isolate populations of a particular species to areas with environments suitable for that species' biotic and reproductive requirements. Restriction of gene flow between subdivided populations by various physical and/or biological barriers acting in both time and space has the potential to generate varying degrees of genetic population structure with a geographical component. Once subdivided, populations may experience intermittent reconnection via migration, or gene flow.

MtDNA Phylogeography of Freshwater Vertebrates

There are numerous recent studies that have utilized mtDNA to characterize the genetic variation and phylogeography of cyprinid populations, as well as of other freshwater vertebrates. On the European continent, Alves et al. (2001) detected population structure with a geographic component in the Iberian cyprinid *Anaecypris hispanica* among nine Spanish rivers utilizing cytochrome *b* and control region mtDNA molecular markers. Populations in two northern and two southern rivers exhibited large site differences, with the authors suggesting evidence of a population bottleneck resulting in haplotype extinction. Populations in five northern rivers possessed low nucleotide diversity, again suggested to be due to bottleneck events but followed by subsequent range expansion. The two southern populations of *A. hispanica* were found to be monophyletic, with an

isolation event separating these populations inferred by the authors. The seven northern and two southern *A. hispanica* populations were proposed as three distinct evolutionarily significant units, or ESUs. An ESU is comprised of regional assemblages of one or more conspecific populations distinguished by highly distinctive clades, resulting from a long-term evolutionary history separate from other such units (ESU; Ryder, 1986). Three criteria are used to designate taxa as an Evolutionarily Significant Unit, or ESU (Waples, 1995). The first is geographic isolation from other populations. The second criterion is genetic differentiation from other related ESUs as measured with neutral molecular markers, due to restricted gene flow. The third criterion is locally adapted phenotypic traits due to differences in selection.

Stefani et al. (2004) examined twelve populations of the cyprinid *Telestes muticellus* on the Italian peninsula, employing mtDNA cytochrome *b* as a genetic marker. The authors ascertained the phylogenetic relationship among five congener species and the pattern of genetic differentiation for the genus in Italy. Intra-population genetic divergence was found to be low, but distinct allopatric lineages were detected inferring recent vicariance coupled with past demographic regression as a possible cause. High levels of genetic differentiation between populations were attributed to the low vagility of *T. muticellus,* which is restricted to the upper headwaters of river basins and disperses between watersheds via headwater capture. The authors suggest

particular attention be paid to the conservation management of *T. muticellus* due to its restricted range and unique distribution of haplotypes.

Machordom and Doadrio (2001) compared twenty-six species of the cyprinid subgenus *Luciobarbus* by analyzing complete cytochrome *b* sequence data throughout its range in Spain, North Africa, Greece, and the Asian Caucasus. The authors found *Luciobarbus* to have a monophyletic origin, but a closer genetic relationship was detected between North African, Caucasian, and Greek species than was found for North African and Iberian *Luciobarbus.* The authors postulate isolation of the main clades occurred after tectonic separation of the Iberian Peninsula and North Africa 5.5 million years before present. Application of a calibrated cytochrome *b* molecular clock from fossil data suggests further splitting of the subgenus *Luciobarbus* among the North African, Greek, and Caucasian populations occurred through isolation events 4.5 million years before present, subsequent to separation of the Spanish peninsula from Africa. Evidence of recent gene flow between *Luciobarbus* populations of the North African Kabilies Mountains and the southern Iberian Peninsula was inferred, as the Algerian species *B. setivimensis* was found to be genetically similar to Iberian *Luciobarbus*. Genetic similarity between North African and Iberian cyprinids has been documented previously in barbels, suggesting recent gene flow between Spain and North Africa (Zardoya and Doadrio, 1998).

Robalo et al. (2006) also investigated the phylogeny and biogeography of an Iberian cyprinid genus, *Chondrostoma.* One hundred and two individuals of

the two species of *Chondrostoma*, *C. arcasii* and *C. macrolepidotus* were studied. Data analysis revealed five phylogroups, the relationship among which was inconsistent with the traditional dichotomy of two species based on morphological descriptions. Based on application of a molecular clock rate calibrated to accepted divergence rates of cyprinid mtDNA cytochrome *b*, the authors propose species differentiation in *Chondrostoma* predates present hydrogeographic basins and instead dates to the establishment of a Miocene-Pliocene endorheic lake system in the Iberian Peninsula.

Costedoat et al. (2006) examined thirty-one populations of the European dace *Leuciscus leuciscus* using three mtDNA markers (cytochrome b, control region, and 16s rDNA) integrated with allozyme and morphological analysis. Five lineages belonging to two clades were detected, and the relationships between lineages were attributed by the authors to Pliocene and Pleistocene paleoclimatic events involving severe climate oscillation, including glacial advance and subsequent retreat. On the basis of their investigation of genetic population structure that included nested clade analysis, the authors suggested four new sites in Western Europe as paleorefugia for *L. leuciscus* during glacial maxima.

Perdices et al. (2005) analyzed the phylogenetic structure of twenty-eight populations of the Asian cyprinid *Opsariichthys bidens* across its range in China. Five mtDNA cytochrome *b* lineages were detected overall within the Yangtze, Pearl, and Hai He drainages with sequence divergence as large as twenty percent observed between lineages, suggesting previous taxonomic

underestimation at the species level. Drainage-restricted haplotypes with high frequency and moderate nucleotide diversity were inferred by the authors as evidence of historic long-term interruption of gene flow, although the distribution of some common haplotypes suggest a recent connection allowing gene flow across drainages.

North American studies of the phylogeography of freshwater vertebrate taxa have shed light on the population structure and evolution of numerous organisms as well. Lemmon et al. (2007) studied 237 populations of trilling tree frog (genus: *Pseudacris*) across North America utilizing 12S and 16S RNA subunit mtDNA sequence data to test correlation of speciation within the genus to known local geologic and climatic events. The evidence supported major speciation within *Pseudacris* as being due to a Mississippi basin marine embayment established during the Pliocene, with genetic variation in the genus reduced due to Quaternary climate change. *Pseudacris* populations presently occupying formerly glaciated areas exhibit significant evidence of population expansion indicating recent colonization, while in arid western populations low levels of variation within clades and low population structure support the hypothesis that recent aridification is responsible for local extinction of populations that have only recently migrated to those regions.

Pfrender et al. (2004) examined thirteen populations of the Oregon speckled dace *Rhinichthys osculus* occupying five major river drainages. Analysis of mtDNA cytochrome *b* sequence data revealed a geographic

population structure characterized by a high degree of divergence among basins and high genetic diversity within basins, with reciprocally monophyletic (single common ancestral) origin of clades. The high diversity within basins was inferred as likely due to the maintenance of large historic population sizes (i.e., lack of bottlenecks) within each basin. Two highly divergent mtDNA lineages were observed to co-occur within the Klamath basin which the authors suggest may represent either represent two reproductively isolated sympatric taxa, or may be due to a complex series of multiple colonization events within the basin. The authors suggest the speckled dace among the major Oregon river basins be regarded as distinct ESUs based their findings.

Spinks and Schaffer (2005) utilized mtDNA control region and ND4 sequence data to ascertain the genetic relationship between populations of the western pond turtle *Emys marmorata* throughout its range from Washington State to Northern Baja California. Four allopatric clades were detected including a northern clade (San Luis Obispo north to Washington State and including Nevada), a San Joaquin Valley clade extending north of the Tehachapi Mountains and west of the Sierra Nevada Mountains, a Santa Barbara clade including the Santa Ynez and Santa Clara river systems, and a southern clade extending from the southern slopes of the Santa Ynez Mountains and south of the Tehachapi and Transverse Mountains through Mohave to northern Baja California. The authors infer recent (< 20,000 years before present) colonization of the northern clade from southern refugia following the last Pleistocene glacial

maxima, with four species extant in what is currently recognized in the literature as one wide-ranging species *E. marmorata*, although the case for division of the species into at least three distinct subspecies on morphological grounds (Seeliger, 1945) had been previously proposed. The Tehachapi and Transverse Ranges are identified as important barriers for gene flow in *E. marmoratus*.

Phylogeography of Southern California Vertebrates

The identification of the Transverse Range as a major phylogeographical break in California has been noted in a wide range of other studies examining the biogeographic population structure of many animal taxa (reviewed in Calsbeek et al, 2003; Chatzimanolis and Caterino, 2007; Forister et al, 2004; Maldonado et al, 2001; Metcalf et al, 2001; Phillipsen, 2005; Phillipsen and Metcalf, 2009; Rodriguez-Robles et al, 1999, 2001; Sgariglia and Burns, 2003). Chatzimanolis and Caterino (2007) note that additional investigations of Transverse Range species are needed for conservation planning, and stress the importance of this relatively small region (e.g., the eastern Transverse Ranges) for containing relatively numerous cases of isolated lineages of multiple widespread species.

Within Southern California few studies have specifically attempted to formally assess the phylogeography of stream-dwelling vertebrate animals utilizing genetic data (Macey et al, 2001; Spinks and Shaffer, 2005; Kuchta and Tan, 2006; Recuero et al, 2006). Tan and Wake (1995) detected distinct mtDNA lineages present in southern California populations of the California newt *Taricha*

torosa. Macey et al. (2001) also found that southern California populations of the mountain yellow-legged frog *Rana mucosa* constituted mtDNA lineages distinct from elsewhere in the state, however neither of these studies focused specifically on southern California and instead sampled populations throughout the entire range of their respective taxa within the state and consequently both studies possess small sampling size from the region.

Phillipsen and Metcalf (2009; also see Phillipsen, 2005) detected geographic population structure in 46 Southern California populations of the California tree frog *Pseudacris cadaverina* using mtDNA cytochrome *b* sequence data from 221 individuals. The authors found the population structure to be associated with the separation of the populations by watershed and mountain ranges. The observed phylogenetic distribution pattern for *P. cadaverina* was inferred to be due to restricted gene flow between separate populations, allopatric fragmentation of desert populations, and range expansion from refugia during the Pleistocene.

Oakey et al. (2004) examined genetic differentiation within *Rhinichthys osculus* across its range in western North America using restriction fragment length polymorphism (RFLP) analysis. Distinct mtDNA lineages were detected in populations within the Los Angeles Basin region (Santa Ana and San Gabriel Rivers) and were identified as most closely related to middle Colorado River *R. osculus* populations. Sampling size within southern California for this study was small due to its broad range, as it focused on characterizing the phylogeny of *R.*

osculus populations across the entire western region of North America rather than specifically within southern California. However, RFLP analysis has limitations, as it may fail to detect much of the genetic variation in closely related, recently differentiated taxa since only mutations producing or removing restriction sites are discernable (Avise, 2000).

Smith and Dowling (2008) investigated the genetic variation present in populations of *R. osculus* sampled from 29 locations in the Colorado River Basin, connective tributaries of the Colorado River, the Los Angeles basin (San Gabriel River), and the Amargosa River (Death Valley system). Molecular analysis of cytochrome *b* and ND4L mtDNA sequences led the authors to propose that Los Angeles basin speckled dace diverged from their closest sister taxa (Gila River dace) in the Colorado Basin approximately 1.7 – 1.9 million years b.p. (s = 1.18 million y.b.p.) based on the observed maximum likelihood branch length of the L.A. Basin clade in the phylogenetic tree constructed from mtDNA sequence data. High haplotype diversity (3 unique haplotypes) is reported to exist within the Los Angeles River clade, with an estimated time of haplotype divergence between 229 – 41 thousand years b.p. Long branch length is attributed to rapid molecular evolution of the L.A. Basin speckled dace due to adaptation to a low-elevation, warm environment relative to populations inhabiting the Colorado basin. However, due to the lack of a constrained time estimate of divergence the connection between L.A. basin and Colorado basin *R. osculus* is left unsolved.

Clearly, additional genetic studies focusing on Southern California's freshwater vertebrate fauna would add considerably to the understanding of the pattern of genetic diversity and biogeography for this region, widely recognized as a significant world biodiversity hotspot (Myers et al, 1999; Calsbeek et al, 2003). Specifically, ascertaining the limits of the range of the Santa Ana speckled dace *R. osculus* in southern California as well as describing its genetic population structure would serve to clarify evolutionary and taxonomic relationships suggested by others, but not fully described to the level of taxonomic certainty (Cornelius 1969; Mayden and Wiley, 1992; Moyle 2002; Oakey et al, 2004). Understanding the geographic limits to the range as well as the population structure of *Rhinichthys osculus* across southern California would serve to suggest specific conservation strategies for management of the fish, once it is known with certainty whether or not the local subspecies of *R. osculus* known as the Santa Ana speckled dace exists as an evolutionarily significant unit (Waples 1995).

Mitochondrial DNA as a Molecular Marker

Molecular genetic studies rely on the use of suitable molecular markers to infer relationships within and among populations. One useful source of molecular markers for genetic studies is mitochondrial DNA (Birkey et al, 1983). Its ease of isolation due to high copy number (Bogenhagen and Clayton, 1974), dual properties of clonal (Hayashi et al, 1985) and maternal inheritance (Dawid et al,

1972) with few exceptions (Zouros et al, 1992; Magoulas and Zouros, 1993; Hoeh et al, 1991), lack of recombination during replication (Dawid et al, 1972) with few exceptions (reviewed by Rokas et al, 2003), and high relative rate of mutation compared to nuclear DNA and thus high level of intraspecific polymorphism has been demonstrated to be particularly relevant in this regard (Moore, 1987).

Mitochondrial DNA: Structure and Function in Vertebrates

The mitochondria is an organelle found in the cytoplasm of eukaryotic organisms which carries out the reactions of oxidative phosphorylation, and is capable of self-replication independent of cellular replication (Clayton, 2000). Self-replication is possible because the mitochondrion possesses its own DNA within the inner mitochondrial membrane space (cristae), typically 5 to 10 DNA copies per mitochondrion (Scheffler, 1999). Mitochondrial DNA (mtDNA) is circular, double-stranded in most cases, and has a structure which is very stable with respect to arrangement and order of genes (Figure 3, appendix A). MtDNA typically contains 37 genes in most multicellular animals and some protozoans, including 22 tRNAs with some exceptions (Cantatore et al, 1987), 13 mRNAs coding for polypeptide chains involved in ATP synthesis and the electron transport chain, and 2 rRNAs. There is also a noncoding sequence of approximately 1kb length called the control region or D-loop, containing the origin of replication and sequences that regulate transcription of the mtDNA genome

(Wilson et al, 1985). Animal mtDNA is typically similar in size across a wide range of taxa with a size of approximately 16kb in length. Animal mtDNA has an overall small size due to the lack of introns and intergenic sequences on the molecule. In contrast, plant mitochondrial DNA is observed to be much larger, in part due to the frequent presence of introns as well as intergenic sequences between genes (Scheffler, 1999).

Structure and Evolution of the Mitochondrial Control Region

The major non-coding control region of mitochondrial DNA is one such useful molecular locus as described above. Mitochondrial DNA has an average rate of divergence of approximately 2% per million years among taxa, a rate that exceeds the observed nuclear DNA divergence rate among the same taxa (Wilson et al, 1985). The rate of control region evolution is reported as typically even more rapid (McMillan and Palumbi, 1997; Tang et al, 2006; Zhu et al, 1994), although in some taxa this is not always the case (Ruokonen and Kvist, 2002; Zhang et al, 1995). Because of this relatively high divergence rate, the control region is useful as a marker for detecting relatively recent evolutionary changes in populations (Dawson et al, 2001; Avise 2000; Baker and Marshall, 1997; Stewart and Baker, 1994; Quinn, 1992).

The major function of the mtDNA control region is as the origin of replication for the heavy strand (O_H) of the mtDNA molecule, as well as containing all known promoters for transcription of mtDNA. The control region for

a variety of vertebrate taxa has been described as consisting of three domains known as Domains I, II, and III (Brown et al., 1986) (see Figure 4, appendix A). The control region of mtDNA is known to be highly A+T rich in insects (Zhang and Hewitt, 1997) and other invertebrates (Flot and Tillier, 2007), but to a generally lesser degree among vertebrates depending on the taxa in question (Hoelzel et al, 1991; Sbisa et al, 1997; Ruokonen and Kvist, 2002; Ray and Densmore, 2002). Within the 3' domain (also known as domain I) are regions termed termination-associated sequences (TAS; Clayton, 1984) that are believed to be associated with termination of the replication cycle of mtDNA. The 5' domain (also known as domain III) contains the origin of replication for the heavy strand (Ori-H) and three regions known as conserved sequence blocks 1, 2, and 3 (CSB-1, 2, and 3) which are believed to act as promoters for initiation of heavy strand replication (Doda et al, 1981; Saccone et al, 1991). CSB-1, 2, and 3 are known to be conserved across a wide range of vertebrate animal taxa (Walberg and Clayton, 1981). The central domain (known as domain II) also contains CSBs involved in heavy strand replication, initiated via a triple stranded displacement loop structure (d-loop; Clayton, 1984) (see Figure 4). Hoelzel et al. (1991) reported the existence of a possible open reading frame (ORF) of more than 100 base pairs in the central domain of three cetacean species (killer whale, minke whale, and Commerson's dolphin). ORFs have been noted in the control regions of other vertebrates (Saccone et al., 1987) as well as for invertebrates (Flot and Tillier, 2007) although their function, if any, remains unknown at

present. Domain II is the most conserved sequence area in the control region of mtDNA (Saccone et al, 1991; Sbisa et al., 1997). Outside of CSB regions there is considerable sequence variation within the control region, even among closely related taxa (Lee et al., 1995). Sbisa et al. (1997) examined control region sequences from ten different mammalian orders. The authors proposed that ETAS1 and ETAS2 (extended TAS) have functional roles in the regulation of transcription, while CSB1 is distinct from CSB2 and CSB3 in that it is common and apparently essential to all mammalian mtDNA genomes. Different functional roles for the conserved elements, CSBs, and TASs would explain differential conservation in the evolution of mtDNA control region. Lee et al. (1995) have noted the similarity of fish CSBs to mammalian CSBs in both structure and arrangement.

Length variation due to tandem repeats is common within mtDNA control sequences of many taxa (Lee et al, 1995; Stepien and Kocher, 1997) and is probably due to secondary structure formation within the mtDNA molecule during replication (Fresco and Alberts, 1960), since gene duplication due to genetic recombination is normally absent (Dawid et al, 1972), although mtDNA recombination has been documented in nematodes (Lunt and Hyman, 1997). However, some vertebrate species are known to exhibit little to no sequence variability within the mitochondrial control region (Walker et al, 1998; Baker and Marshall, 1997). In the absence of recombination events, slipped-strand mispairing (SSM) has been proposed as a primary cause for additions and/or

deletions to the mtDNA control region sequence (Broughton and Dowling, 1997; Zhang and Hewitt, 1997). It has been reported that nearly the entire control region of the self-fertilizing fish *Rivulus marmoratus* has been duplicated within the mtDNA genome of that species (Lee et al, 2001). Guo et al. (2003) mapped the control region of cyprinid carp and report the existence of ETAS in domain I, three central CSBs (CSB-F, CSB-E, and CSB-D)in domain II, and conserved sequence blocks CSB-1, CSB-2, and CSB-3 in domain III similar to those previously reported for other vertebrates. To date, no attempt to describe the structural characteristics of the control region of the genus *Rhinichthys* has been attempted.

The observed rate of mtDNA evolution is approximately 5 – 10 times faster than in single-copy nuclear DNA (scnDNA) in primates, and is generally faster for most vertebrates (Moritz et al, 1987; Schneider and Excoffier, 1999; Nachman and Crowell, 2000). The relatively fast rate of evolution seen in mtDNA compared to scnDNA is probably in part due to mutational restraint on scnDNA, combined with inefficiency of mtDNA relative to the nucleus for repair of replication errors and DNA damage. Since mtDNA does not code for proteins directly involved in its own replication and transcription, less replication accuracy in a small system coding for only 13 proteins would be expected. Loss of mtDNA gene function is probably due to translocation of mitochondrial genes to the nuclear genome, which has been described (Schneider-Broussard and Neigel, 1997). There is evidence supporting the hypothesis that mutational restraint is

lower for mtDNA than for scnDNA. It has been reported that the tRNAs and rRNAs of the mitochondrial genome evolve up to approximately 10 times faster than their nuclear DNA counterparts (Wilson et al, 1985).

Multiple studies (Walberg and Clayton, 1981; Ruokonen and Kvist, 2002; Guo et al., 2003; Tang et al., 2006) report that within the mitochondrial genome itself the rate of mutation of the non-coding control region generally exceeds that of the coding regions, although for closely related taxa of less than ten percent sequence divergence the control region may evolve at a slightly slower rate than coding regions such as *cytochrome b* (Tang et al., 2006). The observed rate of mtDNA control region evolution has been described as up to 43 times more rapid than in other mtDNA loci, including coding sequences such as cytochrome *b* in fish, with a strong transition/transversion (ti/tv) bias (McMillen and Palumbi, 1997), and is generally observed to be rapid for most vertebrates (Aquadro and Greenberg, 1983; Moritz et al., 1987). MtDNA control region evolutionary rates in freshwater fishes have been reported to vary within the range of 3 to 10% change per million years (Stepien and Faber, 1998; Tadao et al., 2001; Takahashi et al., 2001; Yokoyama and Goto, 2002). Evolution rates of DNA sequence within the control region have been shown to vary, with a slower rate of evolution observed in the central domain (domain II), compared to faster evolutionary rates in more variable control-region sectors containing termination-associated sequences (TAS) in domain I and conserved sequence blocks (CSB) in domain III (Hoelzel et al, 1991; Saccone et al, 1991; Sbisa et al, 1997;).

The type of evolutionary change mtDNA is usually observed to undergo is single base substitution (Moritz et al, 1987). Substitutions may occur as either transitions (purine-purine or pyrimidine-pyrimidine substitutions, symbolized as "ti") or transversions (purine-pyrimidine or pyrimidine-purine substitutions, symbolized as "tv"). Hoelzel et al (1991) reported nucleotide substitution rates varied within the control region in a comparison of three cetacean species, with the substitution rate occurring in domains I and III up to 5 fold faster than in domain II, the central conserved region. A similar pattern was reported for Crocodilians (Ray and Densmore, 2002). Guo et al (2003) reported differential substitution rates among the three domains of cyprinid control region mtDNA, with 66.91% of nucleotide positions found to be variable in domain I, 40.00% of positions variable in domain III, and 34.06% of positions variable in domain II in a study comparing seven cyprinid species. In birds, control region nucleotide variation in domains I and III exceed the variation observed in domain II, with the domain of highest variability being genus-dependent (Ruokonen and Kvist, 2002). A thirty-fold bias for transitions (ti) over transversions (tv) where nucleotide substitutions do occur has been reported for the human mtDNA control region (Aquadro and Greenberg, 1983). Zhu et al. (1994) reported that the ratio of ti/tv in the control region was close to 2.0 for closely related freshwater fish taxa (0-5% sequence divergence as measured by cytochrome b divergence), and approaches unity for more distantly related taxa within the rainbow fish genus *Melanotaenia,* with tv actually in excess of ti for highly

divergent taxa. The authors attribute this pattern to saturation of ti due to high A+T content in the control region, as has been observed in *Drosophila* (Wolstenholme and Clary, 1985; DeSalle et al., 1987) and *Cypriniformes* (Guo et al, 2003) as well. Ruokonen and Kvist (2002) report control region sequences in birds become saturated at a ti/tv ratio of approximately 1 when avian species reach 10% pair wise divergence, but at low pair wise divergence of avian species within a genus the ti/tv ratio is higher and may be as high as 40. Transition bias is reported present in the central conserved region of cetaceans, occurring in domain II at levels higher than in comparison to domains I and III (Hoelzel et al, 1991).

Other mtDNA evolutionary changes observed within the control region include length mutations occurring through the processes of tandem duplication or deletion of sequences (Moritz et al, 1987). Lee et al. (2001) report the presence of two nearly identical control regions within the mtDNA genome of the self-fertilizing fish *Rivulus marmoratus* due to a tandem duplication event. Sbisa et al (1997) report domain III to be the most frequent location within the mammalian control region of repetitive sequence motifs/tandem repeats, in a systematic comparison of twenty-seven mammalian genera in which twelve of the twenty-seven mammalian genera were observed to exhibit tandem sequences within domain III. Within the *Crocodylidae* all tandem repeats were observed to occur in domain III as well (Ray and Densmore, 2002). Domain I, however, appears to be the site of all mtDNA control region tandem repeats in

the Pike-Perches genus *Stizostedion* (Faber and Stepien, 1998). Nucleotide substitutions occurring within tandem repeats of control region sequences are reported as useful in reconstructing molecular phylogenies for various taxa including *Ursid* bears (Hoelzel et al., 1994), *Acipenserid* sturgeon fish (Brown et al., 1996), and in the minnow fish *Cyprinella spiloptera* (Broughton and Dowling, 1997).

Because of relatively high divergence rate, mtDNA control region DNA is useful for detecting relatively recent evolutionary divergence between populations of a species. DNA sequencing of selected regions of mtDNA using PCR-mediated amplification has become a common method of choice at present to utilize the mtDNA control region as a molecular marker in evolutionary studies. DNA sequence data is useful in testing and rooting genetic trees for mtDNA evolution (Avise, 2000). Brown et al (1986) cite the rapid rate of mtDNA evolution as support for its neutrality as a molecular marker, but Fry (1999) makes the case that based on neutrality tests using data taken from fourteen molecular studies, mtDNA may follow the mildly deleterious model, or nearly neutral model of evolution. Ballard and Kreitman (1995) state that although genetic drift may ultimately drive mtDNA evolution, if selection does occur on any portion of mtDNA it would influence polymorphism in the molecule in a population as a whole, due to the effect of genetic hitchhiking. Certain knowledge of the neutrality of mtDNA must be ascertained to conduct analyses of genetic distances and molecular clock estimations (Avise et al, 1987; Avise, 2000). Tajima (1989)

describes the statistical method by which neutral molecular evolution may be tested by comparison of the number of segregating sites to the average number of nucleotide differences estimated from pairwise comparison. But regardless of its neutrality, mtDNA has been successfully employed in numerous population level analyses including determination of matrilineal lineages within populations, detecting historical events such as bottlenecks or migrations, and resolving phylogenetic relationships between closely related taxa (Moritz et al, 1987; Avise, 1998; reviewed in Avise, 2000).

To date, no description of the structure or evolution of the mtDNA control region of the cyprinid genus *Rhinichthys* has been reported in the literature. Ascertaining the structure and relative rates of evolution within the various domains and associated sequences of the control region of *Rhinichthys* has direct application toward phylogenetic analysis of the taxa.

Population Genetics

Population genetics seeks to characterize and evaluate the genetic variation within and among subpopulations. F-statistics (Wright, 1931), are calculations which quantify the degree of genetic variation present within and among related subpopulations. The measurement of the degree of genetic differentiation among subpopulations constituting a metapopulation is known as the fixation index, or F_{ST}. The fixation index provides a measurement of the proportional reduction in variation of the metapopulation due to differentiation

among subpopulations, relative to the variation expected with no population subdivision (Conner and Hartl, 2004). Values for F_{ST} range from a minimum of zero to a maximum of one, with increasing value denoting greater genetic differentiation among subpopulations. F-statistics may also be determined at other hierarchical population levels, including estimation of the reduction of variation of individuals within subpopulations (F_{IS}), and for individuals relative to the total metapopulation (F_{IT}). Statistical analogs of F_{ST} developed for analysis of DNA molecular sequence data include N_{ST} (Lynch and Crease, 1990), K_{ST} (Hudson et al., 1992), and Φ_{ST} (Excoffier et al., 1992). Analysis of molecular variance (AMOVA) utilizes the molecular-based analogs of F-statistics to test for the existence of significant population structure. AMOVA performs nested analysis of variance (ANOVA) to compare the degree of genetic differentiation in hierarchically structured populations (Excoffier et al., 1992).

The geographic range occupied by a particular species is a function of its ecological niche, environmental conditions influencing its potential to colonize, and the existence of geographic barriers inhibiting species dispersal (Lomolino et al, 2006). Stream-dwelling vertebrate fish such as *Rhinichthys osculus* occupy habitat that is highly discontinuous across a geographic range due to the available distribution of suitable freshwater habitat required for their metabolic and reproductive needs. Since adjacent freshwater drainages presently disconnected may once have had historical connections and subsequent dissociations, the fish that inhabit them would be expected to have residual

corresponding genetic population structures (Avise, 2000). Statistical tests which are used to infer past population expansion or contraction such as Fu's F_s (Fu, 1997), Tajima's D (Tajima, 1989), and mismatch distribution analysis (Schneider and Excoffier, 1999) can offer insight into the historical demography of freshwater fish such as *R. osculus*.

Evolutionary Forces and Population Differentiation

It is generally recognized that the four processes governing genetic variation between fully or partially isolated populations are mutation, genetic drift, natural selection, and gene flow (Conner and Hartl, 2004). Substitutional mutation involves genetic changes in the four nucleotide bases comprising an individual organism's DNA primarily during DNA replication through the processes of base substitution, deletion, insertion, or inversion. The rate at which mutation occurs is normally slow enough that mutation alone does not significantly alter allele frequencies in natural populations over short timescales. Rather, mutation produces novel genetic variation within a population that the forces of selection or drift may then act upon. Natural selection (Darwin, 1859) is the genetic change a particular population undergoes as individuals within that population express those alleles that enhance survival and reproductive fitness at higher frequency. Genetic drift is the random change in frequency of alleles within a finite population occurring in the absence of natural selection, or when selection is weaker than drift in small populations (Wright, 1931). The process of

gene flow, or migration, tends to homogenize genetically separated populations and thus opposes the processes of natural selection and genetic drift. Geographic isolation tends to facilitate genetic divergence between populations by increasing both genetic drift in smaller populations and natural selection in different habitats, while decreasing gene flow. Genetic drift can assume the dominant role shaping genetic variation in small or recently founded populations (Futuyma, 2005). A new population may begin from a few colonizing individuals possessing a small, random sample of alleles from an ancestral population. This creates one type of population bottleneck, also known as the “founder effect” (Mayr, 1942). Genetic drift in newly founded, small, slow-growing populations tends to alter allele frequencies and reduce genetic variation as compared to the ancestral population (Templeton, 1980; Carson and Templeton, 1984).

Given that Southern California *Rhinichthys osculus* subpopulations inhabit presently isolated watersheds that may have experienced intermittent interconnections during pluvial periods of the Pleistocene 155 – 60 thousand years before present (Smith and Dowling, 2008; Li et al, 2004; Cornelius, 1969; Miller, 1946), it is presumed migration would have played a major role in shaping the evolution of the speckled dace locally, with local drift contributing to genetic differentiation during times of isolation marked by a shift to a more arid climate.

Genetic Structure of Populations and Gene Flow

Intraspecific gene flow between vertebrate fish populations in separate watersheds is normally limited due to the requirement of water-born migration, and their inability to disperse overland. A number of genetic studies of freshwater fish species utilizing mtDNA have revealed the presence of significant phylogeographic population structure consistent with allopatric rather than sympatric origin (reviewed in Avise, 2000). Climatic shifts toward aridity in a region may have the effect of isolating populations of a particular fish species that were once capable of interbreeding due to historical outlet merger of their respective drainages. Once in isolation, the populations may diverge genetically due to the effects of mutation and genetic drift under the mechanism of vicariant population separation (Fisher, 1930; Rosen, 1978). Meffe and Vrijenhoek (1988) proposed two models predicting the genetic population structure of stream-dwelling desert fish. The first model is known as the Death Valley Model and describes the genetic structure of fish populations in complete isolation from one another with zero gene flow between populations. In this model, genetic diversity due to drift within a population is predicted to be low, while genetic divergence among populations within a species is predicted to be high. The second model of Meffe and Vrijenhoek, known as the Stream Hierarchy Model, predicts a more complex hierarchical genetic population structure resulting from gene flow between populations, which is dependent on the geographic proximity and connectivity between populations. Numerous studies of stream-dwelling fauna

(Bunn and Hughes, 1997; Hughes et al, 2000; Hurwood and Hughes, 2001; Preziosi and Fairburn, 1992; reviewed in Avise, 2000) have described population structures consistent with these models. Collecting and assessing genetic data, which reveal the extent of genetic variation among disconnected populations of a particular species may reveal the degree these populations may have shared past connection, as well as their evolutionary history (Avise, 2000; Pfrender et al, 2004).

Proposed Study

This study proposed to utilize specimens of *Rhinichthys osculus* collected from the San Gabriel, Santa Ana, Santa Maria, San Luis Obispo, and Owens River watersheds in Southern California to determine the extent and pattern of the genetic variation within and among speckled dace populations of Southern California, using the mtDNA control region as a molecular marker. Evolutionary theory predicts that in the absence of selection or migration (gene flow), genetic drift acts to genetically differentiate subpopulations of a species. This study will test the hypothesis that subpopulations of Santa Ana speckled dace have experienced restricted gene flow for periods of time long enough to produce significant genetic differentiation.

Genetic data can also be used to draw inferences regarding the historical biogeography of populations. Vertebrate animal taxa inhabiting Southern California have been subject to a complex series of events shaping their

evolutionary history. Studies (reviewed in Calsbeek et al, 2003) have shown that both climate and geology have played important factors in the initiation and maintenance of population divergence in Southern California fauna. Presently, the distribution of southern populations of *R. osculus* is discontinuous and fragmented. This may not have always been the case, in particular during pluvial periods of the Pleistocene and earlier (Thompson et al, 1993). Higher interconnectivity during pluvial periods between now separated watersheds would have influenced the degree of gene flow in the past, which would presumably have been higher than in the present arid condition of Southern California. Recent studies have identified the Transverse Ranges as biogeographical barriers that have significantly influenced the range and population structures of numerous Southern California flora and fauna (reviewed in Calsbeek et al, 2003; Chatzimanolis et al, 2007; Phillipsen and Metcalf, 2009). Populations of northern and southern California speckled dace separated by the Transverse ranges were considered to be similar enough in the past to warrant using the same name (*R. osculus carringtonii*) for both (Culver and Hubbs, 1917; LaRivers, 1952; Hubbs et al, 1974), although this grouping has been questioned on a morphological (Cornelius, 1969) and molecular (Smith and Dowling, 2008; Oakey, 2004) basis. This study will attempt to determine whether southern populations of speckled dace within the Los Angeles Basin (Sampling locations 1 and 2 of Figure 2) are genetically distinct from northwestern populations (Santa Maria River and San Luis Obispo River, sampling locations 3 and 4 of Figure 2,

Appendix A), and northeastern desert populations (Owens River, sampling location 5 of Figure 2, Appendix A) as has been the pattern described for other species.

The final aim of this study is to describe and analyze the structural characteristics and patterns of sequence evolution in the mtDNA control region of the cyprinid mitochondrial DNA control region. One goal of this part of the study includes mapping the locations within the mtDNA control region for the conserved elements with putative functional significance, including conserved sequence blocks (CSBs) and termination associated sequences (TAS). A second goal is to describe the rate and pattern of molecular evolution within the control region, specifically the types, rates, and distribution of nucleotide substitutions throughout the various portions of mitochondrial control region of *Rhinichthys osculus* and the degree to which selection influences these patterns. This will be accomplished by alignment and comparison of the central domain regions from among the study specimens. Thus it is expected that through the study of evolution of variation in speckled dace populations in Southern California, a greater understanding of the molecular evolution of the mtDNA control region of the speckled dace will be obtained.

CHAPTER TWO

MATERIALS AND METHODS

Data Collection

Individuals belonging to the species *Rhinichthys osculus* were sampled throughout its known range in Southern California, which include 3 geographic regions, 5 watersheds, and 5 mountain ranges (Table 1, appendix B). The populations utilized will be based on forming a representative sample of *R. osculus* throughout its Southern California range, as well as specimen availability. Sampling included individuals from populations of *R. osculus* occupying the major drainages of the Los Angeles basin south of the Transverse Ranges, *R. osculus* populations north of the Transverse Ranges taken from drainages of the central coast range, and *R. osculus* populations inhabiting the Owens Valley east of the Sierra Nevada escarpment. Locations of sampling were varied so as to allow for analysis of genetic variation within and among the major Southern California watersheds. An attempt to collect a minimum of seven specimens from each watershed was made where availability allowed. The watershed drainage sampling locations included individuals collected from the Los Angeles Basin south of the Transverse Range (the watersheds of the Santa Ana River and the San Gabriel River), the central coast range north of the Transverse Ranges (the watersheds of the Santa Maria River and the San Luis

Obispo River), and the eastern Sierra Nevada (the Owens River watershed) (Figure 2, appendix A).

Sampling locations within the Santa Ana watershed included individuals of *R. osculus* taken from the following tributaries of the Santa Ana River: Plunge Creek, City Creek, Twin/Strawberry Creek, Cajon Creek, Lytle Creek, and Indian Creek (Figure 6a, appendix A). Sampling locations within the San Gabriel watershed included the following tributaries of the San Gabriel River: the North and West forks of the San Gabriel River, and Cattle Canyon Creek (Figure 6a, appendix A). Sampling locations within the Santa Maria watershed included the following tributaries of the Santa Maria River: Cuyama River, Manzana Creek, and Davy Brown Creek (Figure 6b, appendix A). Sampling locations within the San Luis Obispo watershed included the following tributaries: Stenner Creek, Brizziolari Creek, and San Luis Obispo Creek (Figure 6b, appendix A). Sampling locations within the Owens River watershed included the following tributaries of the Owens River: Marvin's Marsh and Pine Creek (Figure 6c, appendix A). Control Region sequences obtained from GenBank of speckled dace individuals, or individuals of closely related cyprinid taxa, were used to root all native Southern California populations of *R. osculus* in the phylogenetic analysis.

Molecular Methods

Specimens of *Rhinichthys osculus* utilized in this study were collected through the California Dept. of Fish and Wildlife and the United States Forest

Service under the auspices of Anthony Metcalf, Ph.D., Department of Biology, CSUSB. Table 7 summarizes sampling locations and geographic coordinates for all specimens. The protocol for genetic analysis of the speckled dace utilized in this study is summarized as follows. For each specimen, genomic DNA was extracted from muscle tissue using a DNEasy kit, and then visualized using agarose gel electrophoresis. DNA concentration and purity was determined using UV spectrophotometry. PCR amplification of the mtDNA control region was performed using primers ESTFOR/PHE1R, which amplify a mtDNA marker region of 1200 base pairs including the control region (Figure 4, Appendix A). Efficiency of PCR amplification was determined via agarose gel electrophoresis and visualization of the PCR product. DNA sequencing of cloned PCR product was used. For clone sequencing, the PCR product amplicon containing the control region (1200 bp) was cloned into a plasmid vector. The Invitrogen TOPO TA Cloning Kit was used to perform the cloning reactions. Cloning reactions consisted of 3 mL of PCR product (amplified using the same untailed primers already described), 1 mL salt solution (200 mM NaCl, 10 mM $MgCl_2$), 1 mL sterile dH_2O, and 1 mL of TOPO vector incubated at room temperature for 30 minutes. Incubation was immediately followed by bacterial transformation of Mach1-T1 Competent Cells by addition of 2 mL of TOPO cloning reaction to a vial of freeze-dried Chemically Competent *E. coli* cells, and incubated on ice for 5 minutes. The cells were then heat-shocked for 30 seconds at 42^{o}C, followed by addition of 250 mL of S.O.C. medium and incubation with shaking at 37^{o}C for 1 hour. The

resulting transformed cells were plated on LB+kanamycin plates (50 mg/mL) and incubated overnight at 37^{o}C to select for transformed cells. Ten isolated colonies chosen from the plates were selected for analysis and transferred to LB+Kan nutrient broth culture tubes, then incubated overnight at 37^{o}C. Plasmid DNA was extracted and isolated from the broth cultures using a QIAprep Spin Miniprep Kit. The 10 samples of plasmid DNA were tested for the presence of the control region amplicon (1200 bp) by EcoR1 restriction digestion analysis and via PCR amplification of the insert region. Bi-directional cycle sequencing reactions were performed on at least four positive plasmid samples for cloned samples using M13F and M13R primers, which are complementary to M13 primer sites contained within the TOPO vector DNA flanking the control region insertion site. Internal primers designed specifically for this study were used to obtain double stranded control region sequence for each amplicon (Figure 5; Table 2). A minimum of 4 clone sequences were used to generate a double-stranded consensus sequence for each individual over the entire 1200 bp region sequenced to resolve any PCR or sequencing discrepancies. The sequencing reaction products were sequenced using the LICOR model 4300 automated DNA analyzer in the Metcalf lab. Resulting DNA sequences were aligned and any ambiguous bases resolved using LASERGENE (DNASTAR, Inc.) and MEGALIGN (Li-Cor).

Sequence Analysis

The aligned mtDNA control region consensus sequences were analyzed to determine the extent of genetic variation within and among watersheds. This analysis will reveal what, if any, population structure exists for *Rhinichthys osculus* in Southern California, as well as possible historical biogeographical processes that may have played a role in shaping it. In addition, the structure of the mtDNA control region of *R. osculus* will be described for the first time. Examination of the pattern of nucleotide substitutions, insertions, and deletions within the control region may reveal information relating to the molecular evolution of this portion of the mitochondrial genome. Multiple alignments of the sequenced control regions were performed using the software packages Clustal X (Thompson et al, 1997) and MEGA 6.0 (Tamura et al., 2013) to map the presence and location of genetic variation in the various control region domains.

Genetic variation was determined using basic descriptive statistics such as nucleotide frequency, number of polymorphic sites, haplotype number, substitution type and number, number of transitions (ti) and transversions (tv), and the ratio of ti/tv. A common method of evaluating sequence dissimilarity is through calculation of pairwise genetic distances between haplotypes, which is an estimation of the number of nucleotide substitutions per nucleotide site between two DNA sequences. Genetic distance may be uncorrected or corrected, based on factors such as differing substitution rates for ti versus tv, or multiple substitutions at one site. Descriptive statistics were also determined

using Paup* 4.10b (Swofford 2003), GenAlEx 6.501 (Peakall and Smouse, 2006; 2012), MEGA 6.0 (Tamura et al. 2013), and Modeltest 3.06 (Posada and Crandall, 1998). Determining the proportion of variable nucleotides present, and then applying ANOVA to test for significance of any detected differences may reveal relative variability of the control region domains. A minimum-spanning haplotype network was determined using the program TCS (Clement et al., 2000).

Phylogeography

The evolutionary relationships among related mtDNA haplotypes of *Rhinichthys osculus* were represented as phylogenetic trees and estimated using the following methods: 1) genetic distance by minimum evolution (ME); 2) maximum parsimony (MP); 3) maximum likelihood (ML); and 4) Bayesian analysis.

Distance based ME gene trees were calculated using Paup* 4.0b10 (Swofford, 2003) and Mega 6.0 (Tamura et al., 2013). MP gene trees were calculated through both branch-and-bound where appropriate (guaranteed to determine the minimum-length tree) and heuristic searches with bootstrapping at 2000 replicates, using Paup* 4.0b10 (Swofford, 2003) and Mega 6.0 (Tamura et al., 2013). Maximum likelihood trees with 2000 bootstrap replicates were generated using Mega 6.0 (Tamura et al., 2013). Initial trees in the ML analysis were generated by applying Neighbor-Joining and BioNJ algorithms to a matrix of

pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, then selecting the tree topology with superior log likelihood value. A discrete Gamma distribution (+G) was used to model evolutionary rate differences among sites. ML inference employed the Tamura three-parameter (T92) evolutionary model of DNA substitution, which best fit the molecular data as estimated using Modeltest 3.06 (Posada and Crandall, 1998) and Mega 6.0 (Tamura et al., 2013). Bayesian analysis (See Appendix C for MrBayes input file) was performed using MrBayes 3.0b4 (Huelsenbeck and Ronquist, 2001). The initial tree used in the Bayesian analysis was the ML tree described above. Four chains were run in the Bayesian analysis for 30 million generations, with every 100^{th} generation sampled and a burn in fraction of 0.25 of the sampled generations. Since the Bayesian inference of evolutionary history also utilized the T92 model of molecular evolution, the lset command was edited to lset nst=2, and the rate command was set to ratepr = variable. The aligned haplotype data set analyzed was trimmed into 1009 bp sequences, and partitioned into 4 character sets: a short tRNA sequence, Domain I of the control region; Domain II of the control region; and most of Domain III of the control region (Appendix C). This was done so as to allow for the inclusion of 6 control region *Rhinichthys* sequences of 1009 bp from outside of California acquired from GenBank as outgroups. The six outgroups of dace included in the phylogenetic analysis were selected from GenBank based on the limited availability of control region sequence data for speckled dace, or closely related taxa within the genus

Rhinichthys. The outgroups included in the phylogenetic analysis were one of each of the following specimens: the Umpqua Dace *R. evermannii* (Oregon, U.S.; GenBank accession number gi193081196); the Longnose Dace *R. cataractae* (Oregon, U.S.; GenBank accession number gi225194553); the Leopard Dace *R. falcatus* (British Columbia, Canada; GenBank accession number gi224813783); the Speckled Dace *R. osculus* (British Columbia, Can.; GenBank accession number gi224813784); the Millicoma Dace *R. sp. JM2008* (Oregon, U.S.; GenBank accession number gi193203748),; and the Umatilla Dace *R. umatilla* (British Columbia, Can.; GenBank accession number gi225194554).

Population Genetics

Statistical analogs of F_{ST} developed for analysis of DNA molecular sequence data include N_{ST} (Lynch and Crease, 1990), K_{st} (Hudson et al., 1992), and Φ_{ST} (Excoffier et al., 1992). Analysis of molecular variance (AMOVA) utilizes the molecular-based analogs of F-statistics to test for the existence of significant population structure. AMOVA performs nested analysis of variance (ANOVA) to compare the degree of genetic differentiation in hierarchically structured populations (Excoffier et al., 1992). AMOVA was performed using the software programs Arlequin 3.01 (Schneider et al., 2000) and GenAlEx 6.501 (Peakall and Smouse, 2006; 2012). In addition, a Mantel test was used to ascertain the degree of correlation between population pairwise geographic distance and

population pairwise Φ_{ST} values. The results from a Mantel test can be used to infer the presence or absence of isolation by distance among subpopulations.

To provide an estimation of historical demographic change and the direction of that change, mismatch distribution analysis, Tajima's D test, Fu's F_s test, nucleotide diversity (Π), and polymorphic nucleotide sites per nucleotide *q* was employed in the analysis of sequence data. Mismatch distribution analysis (Schneider and Excoffier, 1999) uses the shape of the distribution of nucleotide differences between pairs of DNA sequences (mismatch distribution) under a finite-sites model with mutation rate heterogeneity to infer the presence of past demographic history under a null hypothesis of sudden demographic expansion, for example following a population bottleneck (Schneider and Excoffier, 1999). Populations significantly deviating from the null hypothesis of population expansion (i.e., populations which are demographically stable) will show significant p-values for their mismatch distribution values or their raggedness index (Harpending et al., 1998; Rogers and Harpending, 1992).

Nucleotide diversity Π is the average number of nucleotide differences per site between haplotype sequences, and can indicate direction of population expansion if present. Assuming neutrality, both Π and *q* should be equal. If not, differences may be due to several factors including selection, population expansion, bottlenecks, or mutation rate heterogeneity (Tajima, 1989). The Tajima test of selective neutrality, Tajima's D (Tajima, 1989) was used to determine any significant deviation from the expectation of neutral evolution. This

neutrality test uses a statistical method to compare the number of segregating sites and mean pairwise differences between sequences to test for neutral evolution. Fu's F_s is a statistical method of determining the probability that a nucleotide sequence is selectively neutral, based on the number of observed haplotypes. Departure from the null hypothesis of neutral evolution is detected by these statistics, and may be the result of selective sweeps, population bottlenecks, or demographic expansion (Tajima, 1989; Fu, 1997).

Summary

In summary, a suite of independent analyses examining patterns of phylogeny, population genetics, and historical biogeographical processes were conducted to investigate the evolutionary history of the Southern California Speckled Dace, *Rhinichthys osculus*. An estimation of the influence of landscape on population structure within the *R. osculus* species group in Southern California was determined. Finally, the structural and evolutionary properties of the mtDNA control region of *R. osculus* will be described for the first time.

Speckled dace numbers have declined in Southern California in part due to habitat fragmentation, attributed primarily to human activity including introduction of non-native species, water diversions, urbanization of watersheds, and other human-caused factors (Santa Ana Speckled Dace Recovery Project, 2005). For these reasons the Santa Ana speckled dace was listed as a species

of special concern by the California Department of Fish and Game in 1995 and as a sensitive species by the U.S. Forest Service in 1998.

Following the Southern California fires and flood of 2003-2004, the situation of the Santa Ana speckled dace may have become even more precarious (SAWPA, 2004). A recent U.S. Forest Service survey of the creeks within the Santa Ana watershed all known to formerly contain populations of the speckled dace found only three to still possess dace populations: Cajon Creek, Lytle Creek, and Plunge Creek (SAWPA 2004; Santa Ana Speckled Dace Recovery Project 2005). The speckled dace populations in Twin and City Creeks appear to have been extirpated. Elimination of these speckled dace populations would tend to move the Santa Ana speckled dace toward federal listing, in particular if it could be shown that significant genetic differentiation exists between the populations inhabiting the different drainage systems of the Santa Ana Watershed (Waples 1995). Specific management strategies such as captive breeding programs or species reintroductions, for example, for the remaining watershed populations and drainage areas with the goal of conservation of the species may be suggested once certain knowledge exists regarding the degree of genetic variation among the various remaining Santa Ana Speckled Dace populations (Moritz 1995).

CHAPTER THREE

RESULTS

Sequence Analysis

The entire mitochondrial DNA control region and portions of the flanking mitochondrial genome coding for tRNA-phe, tRNA-pro, and tRNA-thr totaling 1143 base pairs (bp) was successfully sequenced for 74 specimens of *Rhinichthys osculus* representing three geographic regions in California (Southern California, Central Coast, and Eastern Desert). There were 24 unique haplotypes within all populations of *R. osculus* sampled (Table 1). The number of individuals sharing a particular haplotype ranged from 1-19 (Table 1). The control region itself, which is non-coding DNA, accounted for 986 bp of the total 1143 bp of the mtDNA genome sequenced. The overall results for base frequencies within the 1143 bp of mtDNA region sequenced were A = .314, T = .301, C = .234, and G = .150. There were 1051 conserved sites, and 92 polymorphic sites present (Table 3). Of the polymorphic sites observed in the sequence data, 80 (87%) were parsimony-informative. The average uncorrected pairwise (p) distance among all haplotypes was 0.031, with a maximum difference of 0.059. There were 59 transitions and 27 transversions in the control region, with an overall transition-transversion (ti-tv) ratio within the control region itself of 2.19:1 (Table 3).

Structure of the mtDNA Control Region

The control region of *Rhinichthys osculus* is partitioned into three domains as previously described for other taxa (Brown et al., 1986; Guo et al., 2003; Jemt et al., 2015) (Figure 7). In *R. osculus*, Domain I consists of 302 bp, Domain II of 323 bp, and Domain III of 359-361 bp depending on haplotype. Domain I contained 268 conserved sites and 34 (11.3%) polymorphic sites. Of these, 30 sites were parsimony informative. Base frequencies in Domain I were A=0.351, T=0.319, G=0.133, and C=0.197. Domain I contained a total of 20 transitions and 12 transversions, with a ti-tv ratio in Domain I of 1.67:1. Domain II contained 304 conserved sites and 19 (5.9%) polymorphic sites. Of these, 17 sites were parsimony informative. Base frequencies in Domain II were A=0.261, T=0.335, G=0.189, and C=0.214. Domain II contained 13 transitions and 7 transversions, with a ti-tv ratio in Domain II of 1.86:1. Domain III contained 329 conserved sites and 32 (8.9%) polymorphic sites. Of these, 28 sites were parsimony informative. Base frequencies in Domain III were A=0.334, T=0.289, G=0.110, and C=0.267. There were 26 transitions and 8 transversions in Domain III, with a ti-tv ratio in Domain III of 3.25:1. Control region base frequency and substitution data by domain are summarized in Table 3. Total A+T base frequency differed in Domains I (0.670), II (0.596), and III (0.623). A Chi-square test (Table 4) found statistical significance in the frequency of base substitutions among the mtDNA control region domains and tRNAs ($X^2 = 8.535$; $p = 0.036$). Expected values for the number of nucleotide substitutions in each domain were calculated by

multiplying the total number of observed substitutions by the proportion of bases out of the total within each particular domain (Table 4).

The most variable control region domain was Domain I (11.3%), and the domain with the lowest variability was Domain II (5.9%). Domain I has been described in previous studies (Brown et al., 1986; Saccone et al., 1991) as the "hyper-variable" domain of the control region, and this is in agreement with the results of this analysis. Up to 16 short sequence motifs within Domain I, denoted as "termination associated sequences," (TAS) reportedly associated with termination of mtDNA replication, have been described in various studies (Jemt et al., 2015; Ruokonen and Kvist, 2002). A 20 bp TAS motif with the sequence TATGTNTTNNNANCATTNAT, known to be conserved across a wide variety of vertebrates (Jemt et al, 2015) was identified in Domain I (Table 11; Figure 7). The nucleotide motif TACAT has been found to be common within the control region in various vertebrate fish (Guo et al., 2003), although in general are quite non-uniform from taxa to taxa (Ruokonen and Kvist, 2002). In this study, the TACAT motif was identified in more than one location within Domain I in all *R. osculus* haplotypes.

In Domain II, three Conserved Sequence Blocks (CSBs) designated D, E, and F were identified (Table 11; Figure 7), which have been previously found within the central domain of the mtDNA control region of various vertebrates, including fish (Southern et al., 1988; Broughton and Dowling, 1997; Guo et al., 2003). The consensus nucleotide sequence for CSB-D was 20 bp long, with

sequence TTATACTGGCATCTGATTAA. The consensus nucleotide sequence for CSB-E was 19 bp long, with sequence ATGATAGAATCAGGGACAC. The consensus nucleotide sequence for CSB-F was 18 bp long, with sequence ATGTAGTAAGAGACCACC (Table 11; Figure 7).

In Domain III of the control region, three CSBs (CSB-1, -2, and -3) similar to those reported in the literature were identified (Doda et al, 1981; Saccone et al, 1991; Guo et al, 2003). The consensus sequence for CSB-1 was 22 bp long, with sequence AGGTTCATCATTAAAAGACATA. The consensus sequence for CSB-2 was 17 bp long, with sequence CAAACCCCCCTACCCCC. The consensus sequence for CSB-3 was 18 bp long, with sequence TGTCAAACCCCGAAACCA (Table 11). Base frequency and substitution data for each of the three control region domains are summarized in Table 3. A summary of the structure of the mtDNA control region of R. osculus is depicted in Figure 7.

Estimation of the pattern of nucleotide substitution within each control region domain using MODELTEST 3.01 and MEGA 6.0 each yielded a Tamura 3-Parameter plus gamma (T92+G) model of evolution within Domain I, with a gamma shape parameter of 0.29. The T92+G model of evolution also fit the pattern of nucleotide substitution within Domain II and Domain III, with gamma parameters of 0.30 and 0.11, respectively. The overall best-fit model of evolution for the entire mtDNA region sequenced (tRNAs thr/pro + Control Region Domains I/II/III + tRNA phe) was the T92 + G model, with a gamma parameter of 0.21. These results are summarized in Table 5.

Phylogeography and Population Structure

From the complete data set of 74 specimens sampled, a total of 24 unique haplotypes, designated by letters A through X, were identified in *Rhinichthys osculus* in this study (Table 1). The total number of individuals comprising a particular haplotype ranged from 1 to 19. The Santa Ana River watershed populations inhabiting Indian Creek (N=4), City Creek (N=4), Plunge Creek (N=8), Twin/Strawberry Creek (N=12), Cajon Creek (N=13), and Lytle Creek (N=10) contained 12 haplotypes among 57 sampled individuals (Table 1; Figure 9). The San Gabriel River and tributaries (Cattle Canyon Creek, and the West, North, and East Forks of the San Gabriel River) were found to contain 3 haplotypes among 7 sampled individuals. The Santa Ana and San Gabriel Watersheds shared a haplotype common to both, designated haplotype H in the analysis and results (Figure 9). The Santa Maria/Sisquoc Rivers and tributaries (Manzana Creek, Cuyama River, and Davy Brown Creek) contained 2 haplotypes among 6 sampled individuals (Table 1). The Owens River watershed (Pine Creek and Marvin's Marsh, both located in Fish Slough of the upper Owen's watershed) contained 6 unique haplotypes among 6 individuals, meaning each individual sampled represented a unique haplotype (Table 1).

The twenty-four *Rhinichthys osculus* haplotypes identified form three distinct well-supported reciprocally monophyletic clades (Figure 8), based on observed bootstrap values for ME, MP, and ML analysis, and a posteriori probabilities via Bayesian analysis: a monophyletic Southern California clade

consisting of 14 haplotypes found within the Santa Ana and San Gabriel River watersheds (Figure 8); a Central Coast clade consisting of 4 haplotypes found within the Santa Maria and Cuyama Rivers, and the San Luis Obispo Creek watersheds (Figure 8); and an Eastern Desert clade consisting of 6 haplotypes found within the Owens River watershed (Figure 8). The Central Coast and Eastern Desert groups form sister clades, both reciprocally monophyletic with respect to the Southern California clade.

Phylogenetic analysis was determined using previously described methods of distance (ME), maximum parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) methods. Consensus trees of maximum parsimony (MP), minimum evolution (ME), and maximum likelihood (ML) generated using MEGA 6.0 were all congruent and contained similar topologies with each other and with the Bayesian tree (Figure 8). Bayesian inference was performed using the ML tree produced above as a seed tree, with sampling from posterior distribution each 100 generations for 30,000,000 generations using four Markov chains. Sequence data was partitioned in the Bayesian analysis according to the T92+G model of molecular evolution within each respective portion of the control region. The model of molecular substitution best fitting the overall Control Region Domains I-III was the T92+G (Tamura 3-parameter) model of evolution with a gamma shape parameter of 0.21. The T92+G model also best fit the pattern of nucleotide substitution for each of the three domains individually, with gamma parameters of 0.29 in Domain I, 0.30 in Domain II, and 0.11 in Domain III. In the

Bayesian analysis data was partitioned by domain, and also without partitioning. The tree topologies produced using either Bayesian analysis approach were congruent with each other, and with the other methods of phylogeny estimation described above, and so is the only tree depicted (Figure 8). Node values for BI, ML, ME, and MP phylogeny estimation are also shown in Figure 8. The degree of sequence divergence observed between *Rhinichthys osculus* haplotypes ranged from a minimum value of 0.1 percent to a maximum value of 5.8 percent as measured by uncorrected pairwise distances (p-distance) between haplotypes. Bayesian inference of maximum sequence divergence between dace belonging to the Southern California Santa Ana Speckled Dace clade and dace in the Central Coast clade was 7.1%. Maximum BI sequence divergence between Santa Ana Speckled Dace of the Southern clade and dace of the Eastern Desert clade was 7.8%.

Population Genetics

A minimum-spanning haplotype network representing the mutational steps between the14 haplotypes comprising the Southern Clade (Santa Ana/San Gabriel watersheds) was created using TCS and Arlequin, and is depicted in Figure 9. The most common haplotype within the Southern Clade was haplotype G (N=19). Haplotype H was shared between 9 specimens inhabiting the Santa Ana River (N=4) and San Gabriel River (N=5), and was identified by TCS as the

ancestral haplotype (depicted in the rectangular box) in the Southern clade haplotype network (Figure 9).

The haplotype diversity (*h*) within the three clades was all relatively high, with the Eastern Desert clade having the highest value of *h* (1.000) due to the high proportion of haplotypes to individuals, and the Central Coast clade having the lowest value for *h* (0.7333) for a low proportion of haplotypes to individuals. Nucleotide diversity (Π) was low for all three clades, with the Central Coast clade being lowest at Π = 0.000974 (Table 6). This is a result of low genetic differentiation in general within populations comprising a clade. For example, within the Santa Ana watershed, the mean uncorrected pairwise distance (p-distance) within each creek population was as follows: The Indian Creek (IC) population, which was clonal, had a p-distance of zero; City Creek (CC) p-distance was 0.00234; Plunge Creek (PC) was 0.00175; Twin/Strawberry Creek (TWIN) p-distance was 0.00076; Cajon Creek (CAJ) p-distance was 0.00155; and the Lytle Creek (LC) population, which was also clonal, had a p-distance of zero.

Results of neutrality tests including Tajima's D and Fu's F_s, as well as mismatch distribution analysis, indicate that the Southern California clade and the Eastern Desert clade have likely experienced an episode of demographic expansion, possibly following a population bottleneck. Both clades had large negative values for Fu's F_s which were statistically significant (Southern Clade F_s = -5.304 (p= 0.0125); Eastern Desert F_s = -2.696 (p= 0.0300)). Large negative

values for F_s are consistent with the scenario of demographic expansion. The mismatch distribution analysis for the Southern Clade (P= 0.0955; raggedness index = 0.054, p = 0.2835) and the Eastern Desert Clade (P= 0.301; raggedness index = 0.1333, p= 0.509) also supported the inference of demographic expansion, as the data from each did not significantly deviate from presumed models of expansion (Table 6 and Figure 10). The Central Coast clade does not appear to have support for population expansion from mismatch analysis and neutrality tests (P= 0.018; raggedness index = 0.918, p= 0.005). Molecular diversity indices by Region, haplotype and nucleotide diversity, and neutrality tests including Tajima's D, and Fu's F_s were determined using GENALEX and ARLEQUIN and are summarized in Table 6.

Significant spatial genetic population structure for *R. osculus* was found to exist by watershed, mountain range, and geographic region inhabited within California (Southern CA, Central-Coast CA, and Eastern Desert CA). Polymorphic loci are summarized in Table 8. AMOVA (Analysis of Molecular Variance) was used to estimate the genetic variance of *R. osculus* populations in California among populations within regions (Φ_{PR}), among regions (Φ_{RT}), and within populations (Φ_{PT}). Three separate AMOVAs (using Arlequin and GenAlEx) were performed by partitioning *R. osculus* populations by (1) common watersheds (Santa Ana River, San Gabriel River, Santa Maria River, San Luis Obispo Creek, Owen's River; (2) by common mountain ranges (San Jacinto, San Bernardino, San Gabriel, Coast Range, and Sierra Nevada/White Mountains);

and, (3) by common geographic location (Southern California, Central California Coast, Eastern California Desert) (Table 9). Φ_{PT} values among the ten *R. osculus* populations sampled ranged from 0.96 to 0.98 ($p<0.001$) in the three geographic models tested, indicating high levels of population structure among all populations of *R. osculus* (Table 9). The first AMOVA tested for population structure across watersheds and indicated significant population structure by watersheds ($\Phi_{RT} = 0.948$, $p= 0.0127$, $N_m = 0.028$), with 95% of molecular variance attributed to differences among watersheds. The second AMOVA tested for population structure across mountain ranges, and significant structure exists by mountain ranges ($\Phi_{RT} = 0.929$, $p= 0.00098$, $N_m = 0.018$), with 93% of molecular variance attributed to differences among mountain ranges. The third AMOVA design tested for population structure across geographic regions, and once again significant structure exists by region ($\Phi_{RT} = 0.960$, $p= 0.00391$, $N_m = 0.012$), with 96% of molecular variance attributed to differences among geographic regions. The AMOVA analysis summary is shown in Table 9.

Whether populations were grouped by watershed, mountain ranges, or general geographic location within California, the degree of total genetic variation attributed to differences among groups was in excess of 93% ($\Phi_{RT} = 0.948$, 0.929, and 0.960 respectively) in all three models. The model with the highest degree of total genetic variation among *R. osculus* is attributable to genetic variance among geographic regions, at 96% (Table 9). Gene flow N_m among the 3 regions was estimated to be extremely low at 0.012. Principal Coordinates

Analysis (PCoA) confirmed that *R. osculus* exhibits a very high degree of population structure by geographic region (Figure 11). A Mantel Test was performed to assess the presence of isolation by distance. The Mantel analysis (R^2 = 0.6724; p=0.010) revealed significant correlation between population pairwise Φ_{PT} values and geographic distance, supporting isolation by distance of speckled dace populations inhabiting the Southern California Region from the Central Coast and Eastern Desert Regions (Figure 12). A summary of the total percentage of molecular variance observed within populations, among populations, and among geographic regions is shown in Figure 13. Populations also possessed significant genetic structuring within geographic groups, with Φ_{PR} = 0.419, 0.434, and 0.437 for watersheds, mountains, and geographic regions, respectively.

Further population-level genetic analysis within the Southern California Santa Ana Speckled Dace clade was completed by a second AMOVA, with the populations partitioned as follows: 1) by mountain range; 2) by watershed, and; 3) by individual tributaries inhabited by each population. Model 1 (Mountain Range, Figure 14) indicated within the Southern California Region that 21% of molecular variance within speckled dace is accounted for by differences among mountain ranges (Φ_{PT} = 0.212, p= 0.0001; N_m = 1.857). Model 2 (Watershed, Figure 15) indicated that within the Southern California Region 42% of speckled dace molecular variance is accounted for by differences among watersheds (Φ_{PT} = 0.418, p= 0.0001; N_m = 0.696). Model 3 (Tributary, Figure 16) indicated that

within Southern California 45% of speckled dace molecular variance is accounted for by differences among tributaries (Φ_{PT} = 0.448, p= 0.0001; N_m = 0.615). Model 3 accounted for the highest percentage of molecular variance, and therefore differences among speckled dace within Southern California may be most directly attributed to the stream or river they inhabit, and less attributable to the mountain range or watershed inhabited by the dace. Table 9 summarizes the results of AMOVA and Φ-statistic calculations. Population pairwise Φ_{PT} values are summarized in Table 10.

CHAPTER FOUR

DISCUSSION

The purpose of this genetic study of the Santa Ana Speckled Dace *Rhinichthys osculus* was three-fold. The first goal was to characterize the molecular structure of the mtDNA control region of *R. osculus*. This study determined that the molecular structure of the speckled dace mtDNA control region is consistent with studies of other vertebrate animals in the literature.

The second goal of this study was to ascertain the phylogeny of *R. osculus* in Southern California in relation to other speckled dace in California. This research finds that Southern California Speckled Dace comprise a distinct Evolutionarily Significant Unit (ESU) (Ryder, 1986; Waples, 1995) that qualifies the Southern California Speckled Dace as a separate subspecies, if not a separate species in and of itself, from other populations of California fish traditionally identified as *Rhinichthys osculus* on the basis of their morphological characteristics alone.

The third goal of this study was to describe the population genetics of *R. osculus* in Southern California. This study finds that the distribution of speckled dace in Southern California best fits a model of population structure by individual tributary, with episodes of localized population bottlenecks followed by sudden population expansion, most likely linked to climatic variation. This information

suggests conservation and management strategies for the speckled dace populations which remain in existence in Southern California.

Structure of the mtDNA Control Region of Speckled Dace

The mtDNA control region of the Santa Ana Speckled Dace is depicted in Figure 1. The length of the control region was found to be consistently 984 to 986 base pairs (bp) for the 74 specimens of speckled dace sequenced in this study. This length is comparable to control region lengths reported in the literature for other vertebrate fish, such as 1050 bp for dogfish (Delarbre, et. al., 1998), 1006 bp for Atlantic Salmon (Hurst, et.al., 1999), and 928 bp for cyprinid carp and 950 bp for zebrafish (Guo, et. al., 2003); for insects (reviewed by Zhang and Hewitt, 1997), such as 875 bp for grasshopper, and 1029 bp for fruit flies; and in birds (Kvist et.al., 1998), such as 1207 bp for willow tits. Three domains within the speckled dace control region were identified as having characteristics similar to the 3 control region domains ubiquitous to vertebrates (Brown et al., 1986), known as domains I, II, and III (Figure 4 and Figure 7). For speckled dace, Domain I is the most variable control region (11.3% variable sites out of 302 bp), Domain II the most conserved region (5.9% variable sites out of 323 bp), and Domain III intermediate in variability (8.9% variable sites out of 361 bp) (Table 3 and Table 11). This is in accordance with the reported variation in control region domains I-III for animals in general (Tamura and Nei, 1993; Sbisa et al., 1997; Pesole et al., 1999). The differences in numbers of substitutions by Domain was

found to be significantly different by Chi-square analysis (X^2 = 8.535; p = 0.036) (Table 4).

Domain I has been variously described as the “hypervariable” region, or “Extended Termination Associated” region (ETAS) of the mtDNA control region, and is associated with the site of termination of H-strand mtDNA replication (Cheng et. al, 2010; Jemt et. al, 2015). Termination associated sequences, or TAS (Doda et al., 1981), are the putative sites within Domain I where H-strand replication is arrested 95% of the time, after approximately 650 bp from the initiation point, to form a triple-stranded displacement loop commonly referred to as the d-loop (Jemt et. al, 2015). Along with the fastest rate of evolution seen in Domain I (11.3% variable sites), the conserved TAS motif TATGTNTTNNNANCATTNAT found in *R. osculus* and common to a wide range of vertebrates indicates that Domain I in speckled dace follows the same general pattern observed in other vertebrates (Clayton, 1984; Hoelzel et al., 1991; Sbisa et al., 1997). The size of 302 bp seen in *R. osculus* mirrors the sizes reported for other vertebrates which range from 209 bp (*S. araneus*) to 412 bp (*G. glis*; Sbisa et al., 1997). The best-fit model of molecular evolution in Domain I is the Tamura 3-Parameter model (Tamura, 1992) with a gamma shape parameter (T92+G) of 0.29, meaning that the pattern of nucleotide substitution has a G+C bias (low G+C, due to a high A+T content), an unequal rate of transitions to transversions, with nucleotide substitution rates which vary from site to site within Domain I and fitting the gamma shape parameter. The base frequencies and ratios of variable

sites for *R. osculus* within Domain I compared to Domain II (1.92:1) and Domain III (1.27:1) were nearly identical to those reported for other cyprinid fish (Guo et at., 2003) (Table 3).

The patterns of conserved sequences, which are generally known as conserved sequence blocks (CSB), contribute to the relative general rates of base substitutions, and the A+T richness observed in Domains II and III of speckled dace. *R. osculus* shows patterns similar to other vertebrates as well. Walberg and Clayton (1981) designated three conserved sequence blocks within Domain III as CSB-1, CSB-2, and CSB-3. Sbisa et al. (1997) proposed that CSB-1, which is always conserved but to varying degrees among vertebrates, may serve multiple functions including primer generation for mtDNA replication, as well as initiation of mtDNA transcription through binding with mitochondrial transcription factor A (Jemt et al., 2015). For some taxa (primates) CSB-2 and CSB-3 are absent, or partially deleted (Sbisa et al., 1997). Lee et al. (1995) noted the similarity of fish CSBs in structure and function to those in mammals. *R. osculus* mtDNA possesses all three complete CSBs present in Domain III which have been described as common to mammals and birds (Southern et al., 1988; Randi and Lucchini, 1998) (Figures 4 and 7; Table 11). In addition, the more highly conserved Domain II contained 3 CSBs which have been designated CSB-D, CSB-E, and CSB-F (Figures 4 and 7; Table 8) in the literature (Southern et al., 1988; Sbisa et al., 1997; Ruokonen and Kvist, 2002; Guo et al., 2003; Cheng et al., 2010). Since these CSBs are so closely conserved among the populations

which were part of this study, one practical use of knowledge of these conserved sequences is use as primer sites for amplification of control region DNA sequence. Since these conserved elements are relatively uniform throughout the control region (Figure 7; Table 11), these seven short sequences may potentially serve as unique priming sites (TAS, CSB-F,E,D,1,2,and 3) where primers may be designed to allow for PCR amplification of speckled dace mtDNA control region sequence for dace from dace collected from widely different regions. Table 11 summarizes each CSB of *R. osculus*, showing CSB differences (if any) among haplotypes.

The 3 control region domains were determined to possess composition bias similar to that reported for other fish (Guo et al., 2003; Cheng et al., 2010) as well as other vertebrates (Tang et al., 2006). In Domain I the composition bias from highest to lowest base frequency was A>T>C>G; in Domain II T>A>G>C; and in Domain III, the nucleotide bias was A>T>C>G (Table 3). A model of neutral evolution predicts that the substitution rate of a neutral molecular marker will equal its mutation rate (Kimura, 1968). However for a fast-evolving lineage with a high mutation rate, multiple mutations at the same nucleotide site can lead to erroneous interpretations of the phylogenetic signal due to homoplasy, or long-branch attraction (Felsenstein, 1978; 1985; 2004). Rate heterogeneity and nucleotide bias has important implications for estimating divergence between related taxa. A key to understanding the evolution of the mtDNA control region is through analysis of the patterns of nucleotide polymorphisms present. Detailed

knowledge regarding the behavior of mtDNA control region domain of the speckled dace will lead to more accurate estimations of phylogeny using mtDNA sequence data (Felsenstein, 2004). The similarity of the molecular structure and patterns of nucleotide substitution within the control region of the speckled dace mtDNA genome to a wide range of other taxa indicates that the molecular evolution of the control region of this taxa likewise follows a similar pattern.

Phylogeography and Population Structure

The pattern and extent of genetic variation in Santa Ana Speckled Dace demonstrates significant population structure within and among geographic regions and populations. Twenty-four unique mtDNA control region haplotypes of *R. osculus* were identified as inhabiting 5 major drainages in California (Table 1 and Figure 8). The genetic variation of these 24 haplotypes was found to be most strongly associated with the 3 geographic regions of Southern California (14 haplotypes), the Central California Coast (4 haplotypes), and the Eastern California Desert (6 haplotypes), although nearly equally strong association between genetic variation and major watersheds, and between genetic variation and mountain ranges also exists (Table 9 and Figure 13). No shared haplotypes were found among populations sampled from the different regions of Southern California, the Central California Coast, and the Eastern California Desert (Table 1). This indicates an absence, or extremely limited degree, of gene flow between Southern California and these 2 other geographic regions ($Nm \leq .012$). The

Santa Ana Speckled Dace represent an isolated assemblage of R. osculus populations, evolutionarily independent from other regions containing speckled dace. This conclusion supports the finding of Nerkowski (2015) performed using microsatellite analysis of the same populations of speckled dace. Shared haplotypes were found to be common among local populations of speckled dace inhabiting streams in close geographical proximity (Table 1).

Within the Santa Ana River watershed, low levels of gene flow were observed in varying degree among most of the creeks (Table 10). However, pairwise Φ_{PT} values were significant between all Santa Ana watershed creeks except for two creeks which shared haplotypes with City Creek, which had a very small sample size (N = 4): City and Plunge Creeks (Φ_{PT} = 0.211; P = 0.062; Nm = 1.872); and City and Cajon Creeks (Φ_{PT} = 0.158; P = 0.073; Nm = 2.670). Indian Creek dace was the only subpopulation not sharing at least one haplotype with another creek within the Santa Ana watershed. However, Indian Creek is more geographically isolated from the other Santa Ana creeks as it is a tributary of the San Jacinto River, and as such only intermittently connected to the Santa Ana River, and only then during high precipitation years at flood stage via Lake Elsinore and Temescal Canyon Wash (Figure 6a).

Evidence of gene flow between the Santa Ana and San Gabriel River watersheds was found in Twin Creek. Four specimens sampled from Twin Creek (Santa Ana R.) shared the same haplotype (H) as 2 specimens from Cattle Canyon Creek, and 3 specimens from the East, North, and West Forks of the

San Gabriel River, respectively (Table 1; Figure 9). This indicates gene flow has occurred between the San Gabriel River and the Santa Ana River. One possible explanation for the shared haplotype H of this study is that it is a relict haplotype from pluvial times, when the Santa Ana and San Gabriel watersheds were more likely to share interconnection. However, since the Santa Ana River and the San Gabriel River watersheds have not experienced interconnection within historical times, an alternative explanation is recent transplantation of fish from the San Gabriel River to the Santa Ana River, possibly by sport fishermen in the course of stocking creeks with a food supply of small fish for larger stocked game fish such as trout. The practice of stocking small fish to serve as food for larger game fish is known to occur where sport fishing exists (Hubbs and Miller, 1948). Phillipsen and Metcalf (2009) discovered that populations of the California tree frog *Pseudacris cadaverina* inhabiting Lytle Ck., City Ck., and Mill Ck. of the Santa Ana Watershed and populations within the East Fork and West Fork of the San Gabriel River also share a single haplotype. The fact that shared haplotypes between the Santa Ana and San Gabriel watersheds exist for two different freshwater taxa inhabiting the same streams gives support to the first hypothesis, of past hydrological interconnection between the Santa Ana and San Gabriel watersheds and migration of speckled dace between these two watersheds which are now physically separated.

Statistical analysis of population genetic data within the Southern dace clade (which includes the Santa Ana and San Gabriel watersheds) indicates

evidence of population expansion within that clade. Fu's F_S for the Southern California clade as a whole was statistically significant, and indicative of population expansion (Fs = -5.30392; P = 0.0125) (Table 6), and this was supported by the results of mismatch distribution analysis, which also indicated population expansion within the Southern dace clade (Table 6; Figure 10). The model of expansion assumed as the null in the mismatch analysis is one of a single, sudden population increase (Schneider and Excoffier, 1999). This may be explained by 2 different scenarios with regard to the Santa Ana Speckled Dace.

The first possibility is one of a population bottleneck for an existing established population, with rapid population expansion subsequent to the bottleneck as the population recovered. This hypothesis would be consistent with the conditions freshwater vertebrates are exposed to in the present time, and recent geologic past of California's climate, in particular following the end of the Pleistocene (Harden, 2004; Thompson et al, 1993). Since then, as the climate has progressively warmed and tended to arid conditions in general with intermittent periods of high precipitation in intervening years, isolated stream-dwelling vertebrates such as the Speckled Dace have become vulnerable to the effects of drought, fire, and flood and their deleterious effect on local populations (SAWPA, 2004; Santa Ana Speckled Dace Recovery Project 2005).

A second possible explanation for the existence of a population bottleneck in the demographic history of the Speckled Dace is through the founder effect (Mayr, 1942). This type of population bottleneck is the result of a small founding

population of individuals representing random sampling of haplotypes from an ancestral population, migrating into a new region and establishing a new small population, which then undergoes demographic expansion, spatial expansion, or both. If the founder effect were the correct explanation for the inferred population bottleneck for the Santa Ana Speckled Dace, then we would expect to find shared haplotypes, or at the least haplotypes with only a few mutational differences, with an ancestral population in an adjacent geographic region. Since there are no such shared haplotypes observed between Southern California and other geographic regions known to be inhabited by *R. osculus*, and very low gene flow is present as estimated by AMOVA ($N_m = 0.012$), we take this hypothesis as the least likely explanation. Molecular evidence appears to support the first hypothesis, that of a large decrease in size of an established speckled dace population due most likely to environmental or climatic effects, followed by expansion. In addition to the large negative of Fu's F_s statistic and mismatch results, support to this interpretation is seen in the haplotype network for the Southern California dace (Figure 9), in which haplotypes form several star-like patterns, one indication of demographic expansion (Cooke et al., 2012).

The population-level analysis of the Santa Ana Speckled Dace within Southern California indicates that the major factor influencing genetic variation among Santa Ana Speckled Dace is attributed to the stream or creek they inhabit (Figure 16). This is in accord with the Stream Hierarchy Model of Meffe and Vrijenhoek (1988), which predicts a hierarchical genetic population structure

resulting from gene flow between populations, dependent on the geographic proximity and connectivity between populations. Numerous studies of stream-dwelling fauna (reviewed by Avise, 2000) fit this type of population model. From Table 10 we see that the pairwise haploid migration rate Nm is generally higher for populations closer together (upper values in each column), and generally decreases as the geographic distance between populations increases (lower values in each column). For example, City Creek is immediately adjacent to Plunge Creek, Twin Creek, and Cajon Creek. The haploid pairwise N_m values are 1.872, 0.423, and 2.670, respectively. Two of these pairs, City/Plunge Ck., and City/Cajon Ck., have the only non-significant pairwise population Φ_{PT} values due to the high level of gene flow which has occurred between these creeks.

Phylogeographic Analysis

This study was the first investigation of the regional phylogeography of speckled dace (*Rhinichthys osculus*) in Southern California based on mtDNA control region genetic data. Three distinct reciprocally monophyletic clades of speckled dace were identified in this study that correspond to the three geographic regions sampled: A Southern Clade consisting of dace inhabiting the San Gabriel and Santa Ana watersheds; a Central Coast Clade consisting of dace inhabiting the Santa Maria River and its tributaries, and the San Luis Obispo Creek; and an Eastern Desert Clade consisting of dace inhabiting the

Owens River and its tributaries (Figure 2; Figure 5). The phylogenetic analysis indicates that the Southern California Speckled Dace is an Evolutionarily Significant Unit (ESU), significantly genetically distinct from other populations of speckled dace in California.

The sequence difference as estimated by Bayesian inference between the Santa Ana Speckled Dace and the dace inhabiting the Central Coast of California is a maximum of 7.1%. The maximum genetic distance between the Santa Ana Speckled Dace and the Eastern Desert dace population is 7.8% as measured by Bayesian branch lengths. Bayesian estimation would most likely provide a better estimate of the true genetic distance between sister clades than would using uncorrected p-distances, which indicate the maximum genetic distance between the Santa Ana dace and the other populations is 5.8%. This is because the model used in the Bayesian estimation takes into account the likelihood of multiple hits at the same nucleotide position, which would result in an underestimation of the true mutation frequency and sequence evolution between related taxa if measured by uncorrected p-distance (Huelsenbeck and Ronquist, 2001). The Bayesian estimation of the phylogenetic tree and branch lengths was carried out given an a priori model of evolution (in this case, the T92 + G, Tamura 3-parameter model) best fitting the observed control region data for the dace. Within-clade genetic variation of the Santa Ana dace is relatively low in comparison, ranging from 0-1.1%.

Tang et al. (2006) report that for control region (CR) sequences of Cobitoid fishes (Teleostei: Cypriniformes) of the same species but in different populations, CR sequence divergences less than 4.8% for most sequences are expected. However, CR sequence divergences greater than 5.2% are indication of species- or genera-level divergence (Tang et al., 2006). Given that the cyprinid *R. osculus* is also a member of the order Cypriniformes, and that sequence divergence of Santa Ana Speckled Dace compared to other California speckled dace populations exceeds 7%, there is support to designating the Santa Ana Speckled Dace as a distinct species separate from the dace of the Central Coast and the Eastern Desert not only on a genetic basis, but also by morphological consideration as well. Santa Ana Speckled Dace differ from other California dace populations identified as *R. osculus* by having smaller scales, a better-developed frenum on the upper lip, a longer head, and smaller eggs based on the analysis done by Cornelius (1969).

Previous studies (Cornelius, 1969; Oakey, 2004; Smith and Dowling, 2008) have hypothesized that Southern California populations of speckled dace are the result of migration of dace from populations inhabiting the Colorado River Watershed, rather than from dace populations located in northern California. Cornelius (1969) proposed that Central California Coast dace populations were derived from Southern California populations, while the Eastern California Desert (Owens River) populations of dace originated from dace populations to the north, such as those found inhabiting the Klamath River. Since no Colorado River dace

were successfully sequenced in this study, the first hypothesis cannot be tested. However the second hypothesis of Cornelius, that Southern California Dace populations migrated into the Central Coast and therefore are the ancestral population of the Central Coast Clade, is not supported. Rather, this study finds that Southern California dace populations share a common ancestor with both the Central Coast and the Eastern Desert dace, which are slightly more closely related to one another than to the Southern California dace populations (Figure 2; Figure 5). The absence of significant gene flow ($N_m = 0.012$) between the three regions of Inland Southern California, Central California Coast, and Eastern California Desert is one consequence of isolation by distance, as summarized in the results of the Mantel Test depicted in Figure 6 ($R^2 = .672$; $p = 0.001$). This isolation is a result not only of linear geographic distance, but also the existence of physical barriers to migration including separation by intervening mountain ranges, which fragment dace populations into inhabiting separate watersheds associated with the various mountain barriers (Minckley et al., 1986; Moyle, 2002). Central Coast dace populations inhabit watersheds west of the Coast Range Mountains, while Eastern Desert dace populations inhabit watersheds east of the eastern escarpment of the Sierra Nevada Mountain Range. Southern California dace populations inhabit watersheds associated with the San Jacinto, San Bernardino, and San Gabriel Mountain Ranges. Significant results for geographic models testing genetic variation within and among speckled dace populations were found whether populations were partitioned by Watersheds

(Table 9; Φ_{RT} = 0.95; p < 0.001) or by Mountain Ranges (Table 9; Φ_{RT} = 0.93; p < 0.001). These models found nearly the same results as partitioning populations by geographic region (Table 9; 96% of molecular variance attributed to differences among regions; Φ_{RT} = 0.96; p < 0.001). The existence of physical barriers, including mountain ranges and discontinuous waterways, between the Southern California dace populations combine to limit gene flow between these populations, and thus isolate Southern California dace from other California populations. Further sampling of dace in intervening zones between these three regions may help to shed light on the historical development of dace population structure in California.

Santa Ana Speckled Dace

Significant population structure was found to exist among the Southern California Speckled Dace populations comprising the Santa Ana Speckled Dace. Four models of partitioning dace populations, to ascertain the level of dace population structure and gene flow were used: 1) mountain ranges containing the headwaters of tributaries inhabited by dace (Figure 8); 2) watersheds containing the tributaries inhabited by dace (Figure 9); 3) separate tributaries considered as distinct populations (Figure 10), and; 4) isolation by distance (Figure 11). Of these, it was found that the greatest percentage of molecular variance among dace populations is described by the tributary model (Figure 10). Forty-five percent of molecular variance is attributable to differences among tributaries; Φ_{PT}

= 0.448; p = 0.0001; N_m = 0.615). Considering that as climate in Southern California has tended to become more arid following the end of the last ice age, streams that were formerly interconnected during pluvial times have since become more fragmented and isolated from one another. This would contribute to populations of stream-dwelling organisms with low vagility such as dace becoming less likely to migrate, and therefore local populations within streams beginning to evolve in relative isolation through neutral evolution via genetic drift. We see in the data that the larger scale relationships among Speckled Dace populations, fragmentation and isolation by distance and watershed mirrors the population structure seen on the smaller scale, through isolation by distance and by tributary, in the population-level analysis. The same environmental factors are at play shaping both the large-scale and small-scale structure of the Santa Ana Speckled Dace.

Conservation Implications

Three criteria are used to designate taxa as an Evolutionarily Significant Unit, or ESU (Waples, 1995). The first is geographic isolation from other populations. The second criterion is genetic differentiation from other related ESUs as measured with neutral molecular markers, due to restricted gene flow. The third criterion is locally adapted phenotypic traits due to differences in selection. The Santa Ana Speckled Dace fulfills all three of these criteria. It is proposed that its large genetic distance (>7% corrected distance) from related

populations, in combination with low levels of gene flow ($N_m \leq .012$) with those populations and different phenotypic traits (see Cornelius, 1969) when compared to those populations warrant the Santa Ana Speckled Dace recognition minimally as an ESU and subspecies status, if not full status as a separate species from other California populations now identified as *R. osculus*. It is encouraged that policies be enacted to ensure the survival of this unique vertebrate within its natural habitat within Southern California as soon as possible. Within the existing extent of the Santa Ana Speckled Dace habitat, specific refuges should be designated as Aquatic Diversity Management Areas (Moyle, 1995). Surveys similar to the Lytle Creek Survey of 2010 should be undertaken to establish the full range of habitation the Santa Ana Speckled Dace presently occupies, and could potentially reoccupy in future. Finally, management strategies aimed at conservation of the remaining stock of dace may be put into place, with the aim of reintroduction of dace into areas formerly occupied by the dace, but now extirpated due to human activity. The results of this and other studies are resources from which a clear management strategy may be derived (Stephenson and Calcarone, 1999).

APPENDIX A

FIGURES

Figure 1. *Rhinichthys osculus*, the Santa Ana Speckled Dace (Barrett, P. 2003. USFWS).

Figure 2. *Rhinichthys osculus* California Range Map. Areas bordered in dark are known to contain *Rhinichthys osculus* within river drainages of that region. Proposed watershed sampling locations are denoted in red numbered circles. 1=Santa Ana River; 2=San Gabriel River; 3=Santa Maria River; 4=San Luis Obispo Creek; 5=Owens River.

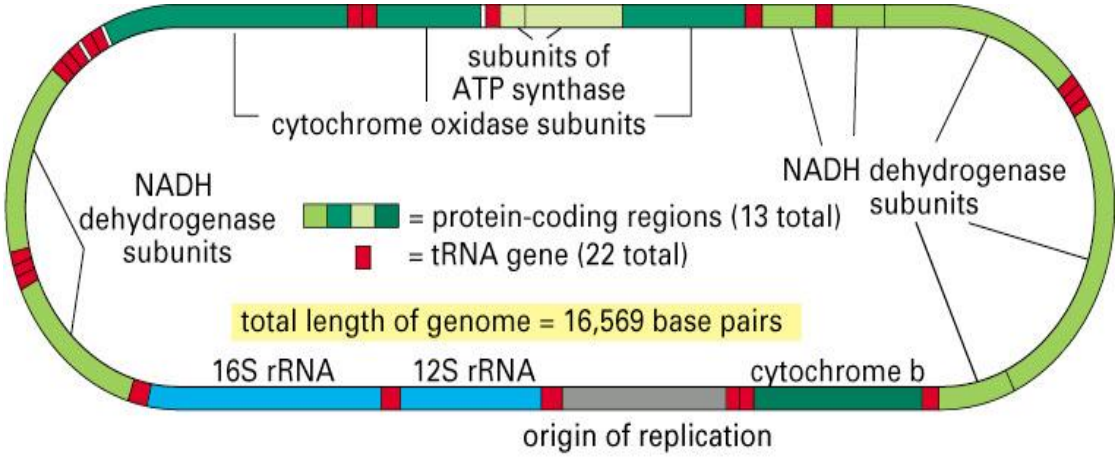


Figure 3. Structure of Vertebrate Mitochondrial DNA. The noncoding mtDNA control region is shaded in gray and contains the origin of replication for both the heavy and light strands of mtDNA (from Molecular Biology of the Cell,4th ed.).

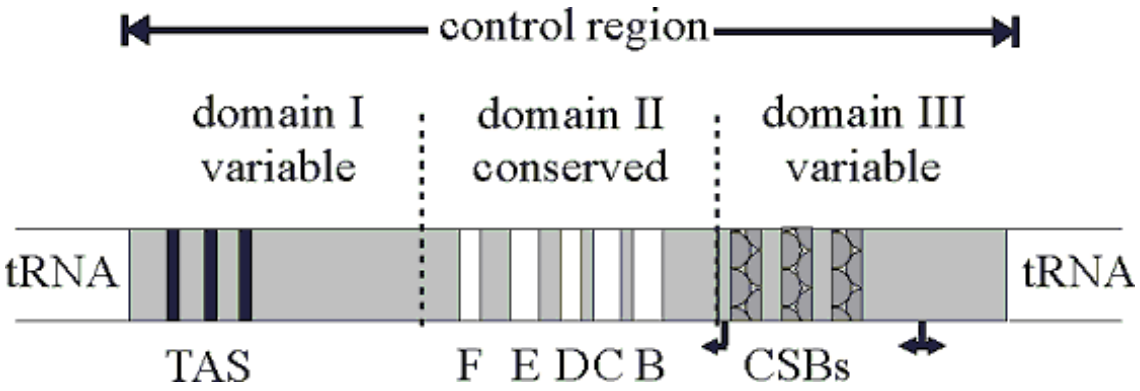


Figure 4. Structure of the Vertebrate Mitochondrial DNA Control Region
The arrows indicate the location of the H-strand replication origin and the bidirectional promoter for L- and H-strand transcription. TAS, termination associated sequence; F through B, conserved sequence boxes (CSB) in the central domain; CSBs, conserved sequence blocks in domain III. (From Ruokonen and Kvist, 2002).

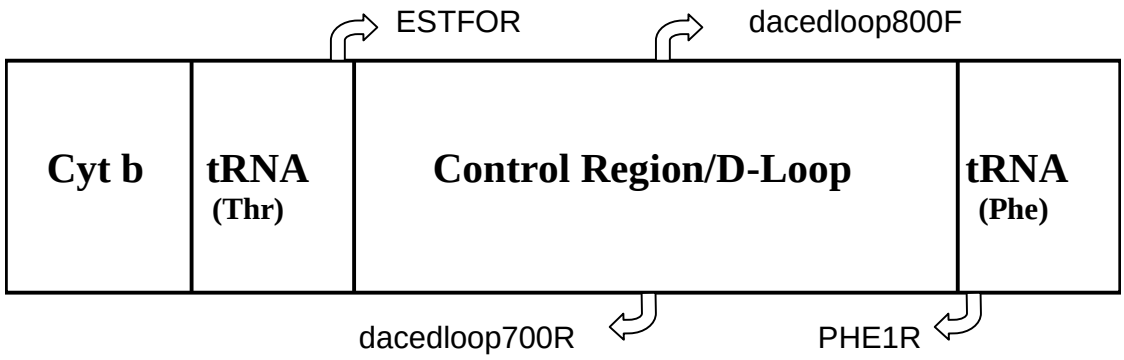


Figure 5. Map of Primer Locations on the Mitochondrial Genome of *Rhinichthys osculus*. Approximate locations of primers used in PCR and sequencing are shown. Forward primers are depicted above and reverse primers below the diagram. See Table 2, Appendix B for primer sequences. Figure is not shown to scale.

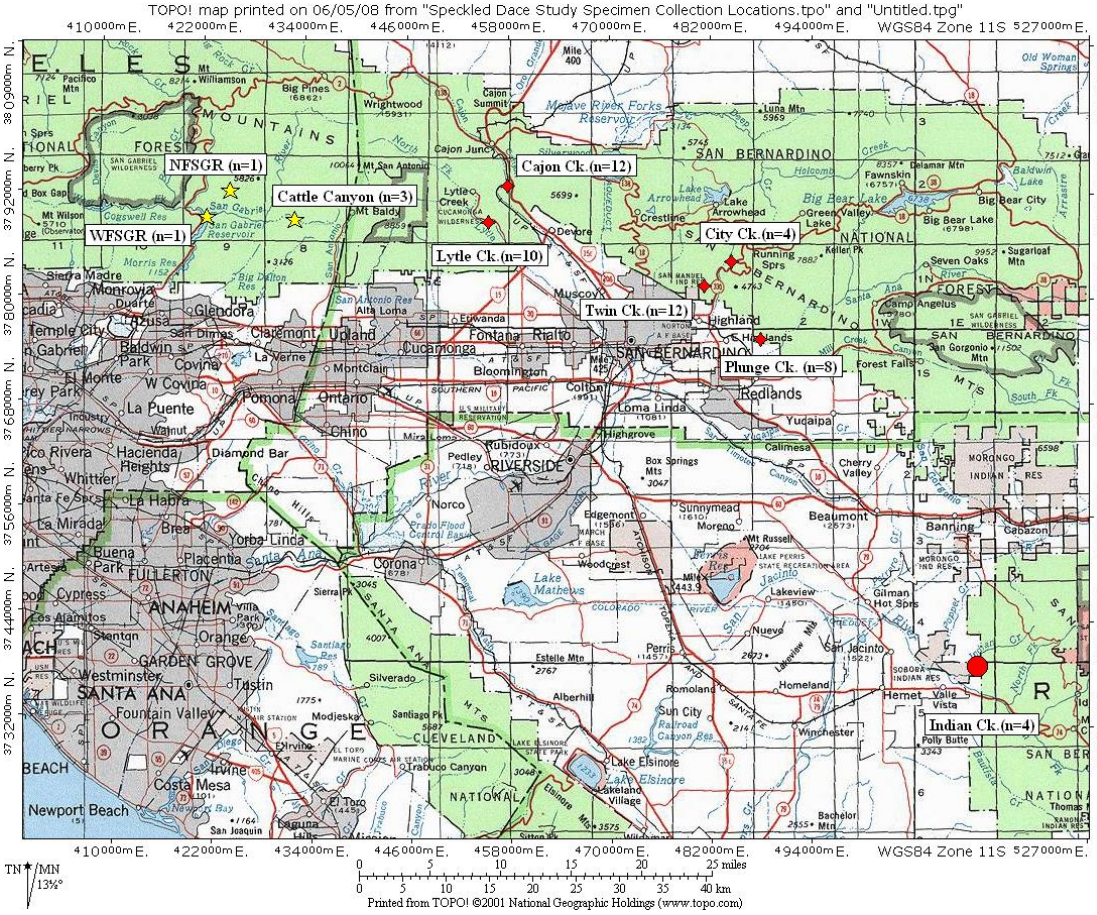


Figure 6a. *Rhinichthys osculus* sampling sites in the Santa Ana River (SAR) and San Gabriel River (SGR) Watersheds. SAR sites = Lytle Creek, Cajon Creek, Twin Creek, City Creek, Plunge Creek, and Indian Creek; SGR sites = North Fork SGR, West Fork SGR, Cattle Canyon Creek.

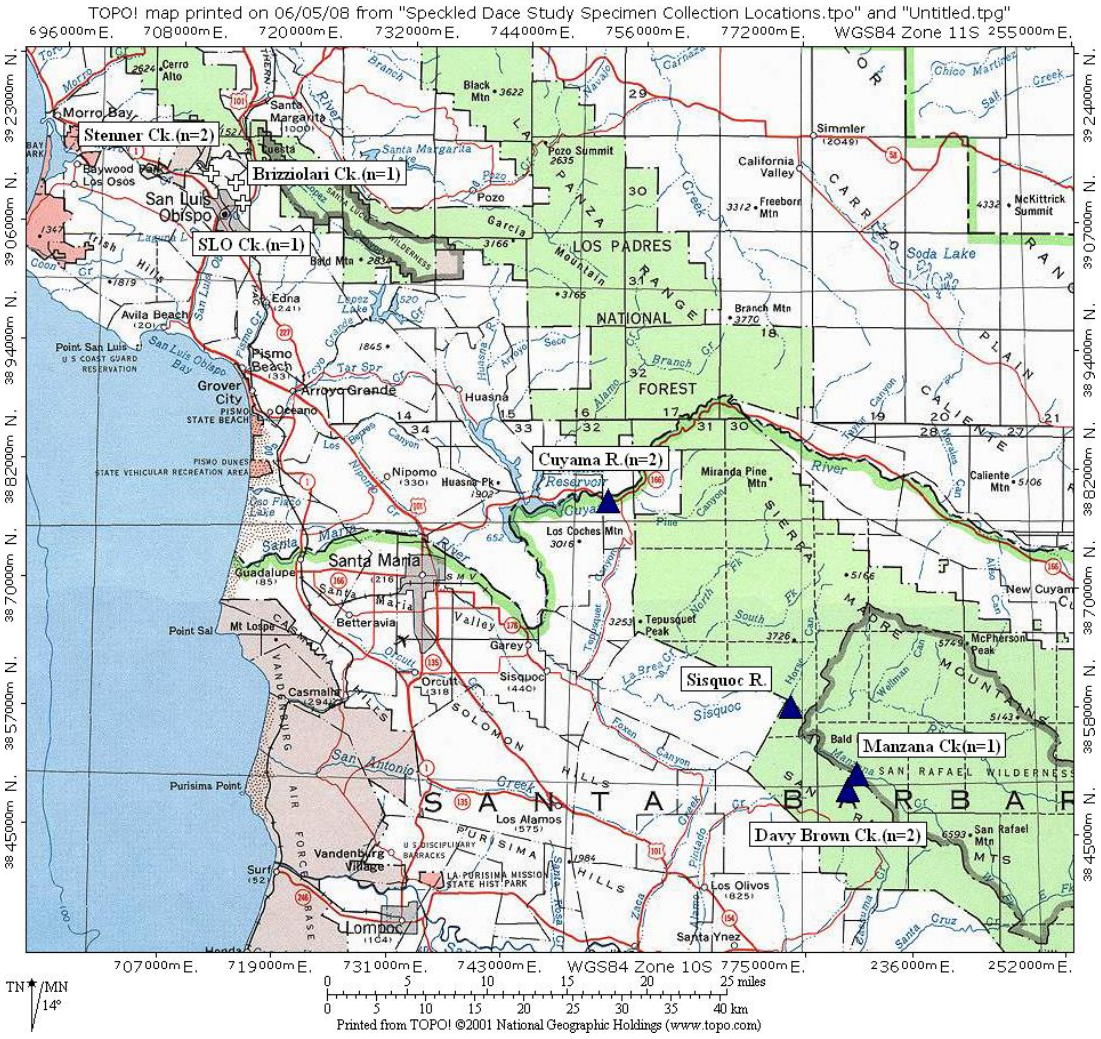


Figure 6b. *Rhinichthys osculus* sampling sites in the San Luis Obispo River (SLO) and Santa Maria River (SMR) Watersheds. SLO sites = San Luis Obispo Creek, Stenner Creek, and Brizziolari Creek; SMR sites = Cuyama River, Manzana Creek, and Davy Brown Creek.

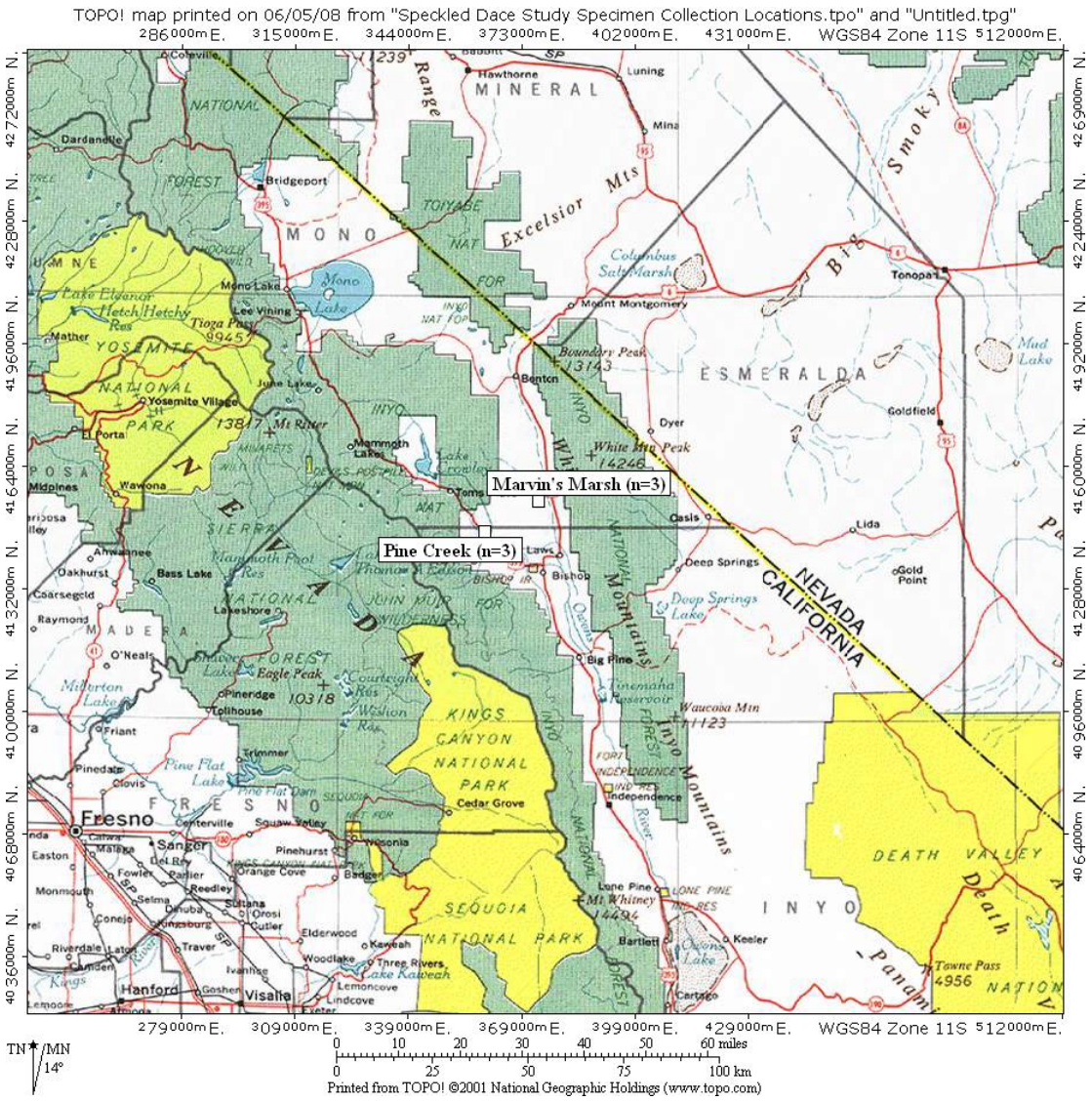
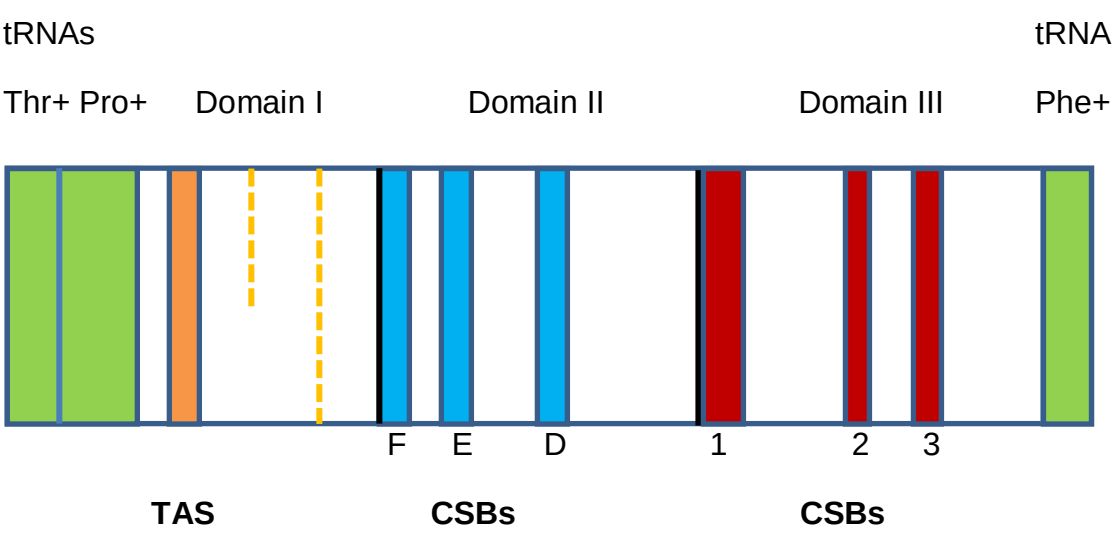


Figure 6c. *Rhinichthys osculus* sampling sites in the Owens River Watershed. Owens River sites = Marvin's Marsh and Pine Creek.

mtDNA Region	base pair begins:	base pair ends:	total length (bp)
tRNA(thr+)	1	48	48
tRNA(pro-)	49	118	70
mtDNA Control Region	119	1104	986
mtDNA CR Domain I	119	420	302
mtDNA CR Domain II	421	743	323
CSB F/Domain II	421	438	18
CSB E/Domain II	466	484	19
CSB D/Domain II	572	591	20
mtDNA CR Domain III	744	1104	361
CSB I/Domain III	744	765	22
CSB II/Domain III	864	880	17
CSB III/Domain III	906	923	18
tRNA(phe+)	1105	1143	39



Termination Associated Sequence (TAS) = TATGTNTTNNNANCATTNAT

TAS motif (TACAT) =

Figure 7. The structure of the mtDNA Control Region of *Rhinichthys osculus*. Green represents tRNAs; orange represents termination associated sequence; blue and red represent conserved sequence blocks in Domains II and III; dotted yellow line represents TAS motif.

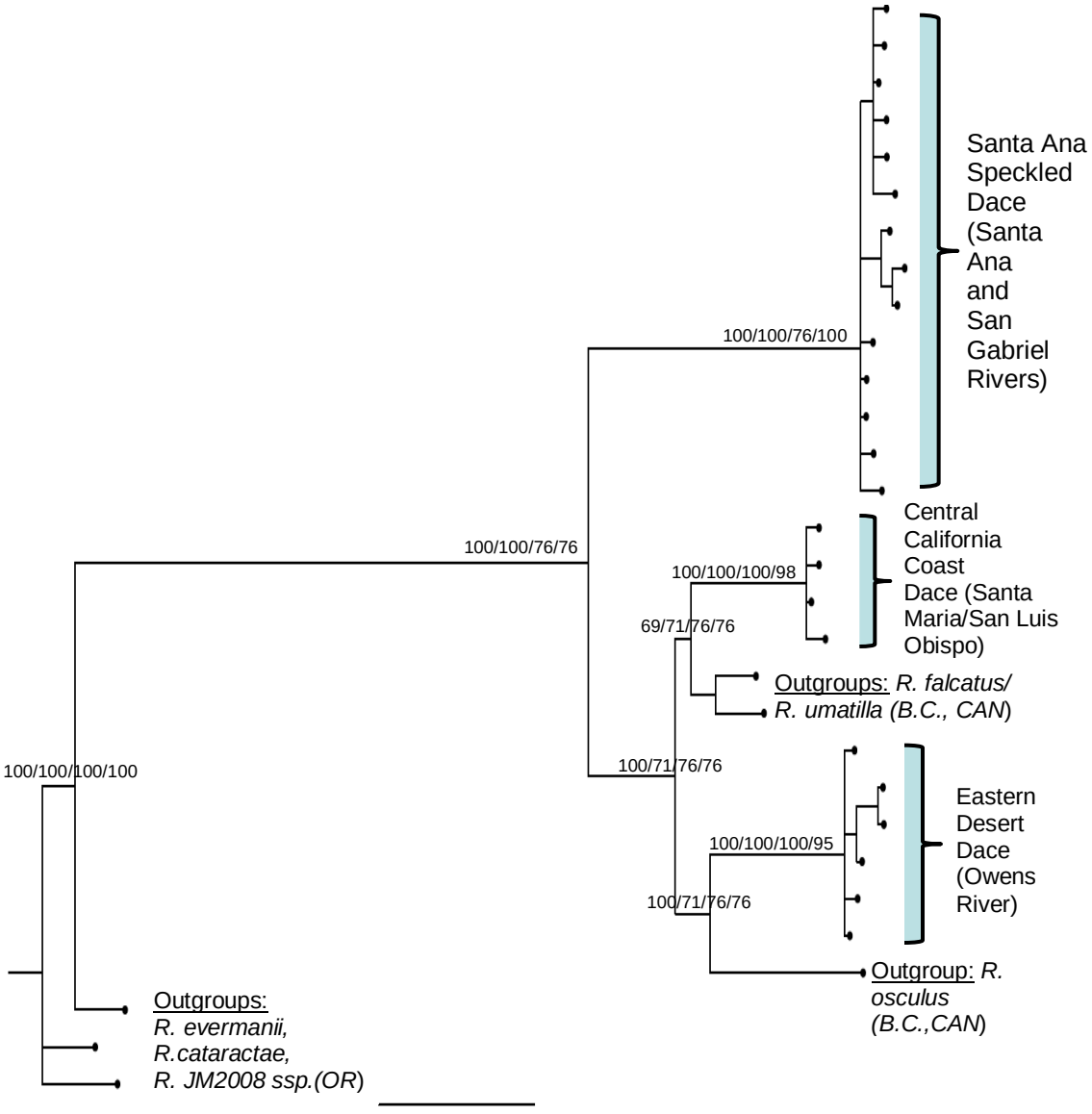


Figure 8.. Bayesian Inference Tree. The evolutionary history was inferred using Bayesian analysis based on the Tamura 3-parameter model. Maximum likelihood (ML), maximum parsimony (MP), and minimum evolution (ME) analysis produced congruent trees (not shown). Genetic distance scale at bottom of figure. Bayesian posterior probabilities, and bootstrap support (2000 replicates) for ML, MP, and ME trees are shown at nodes in the order: Bayes/ML/MP/ME.

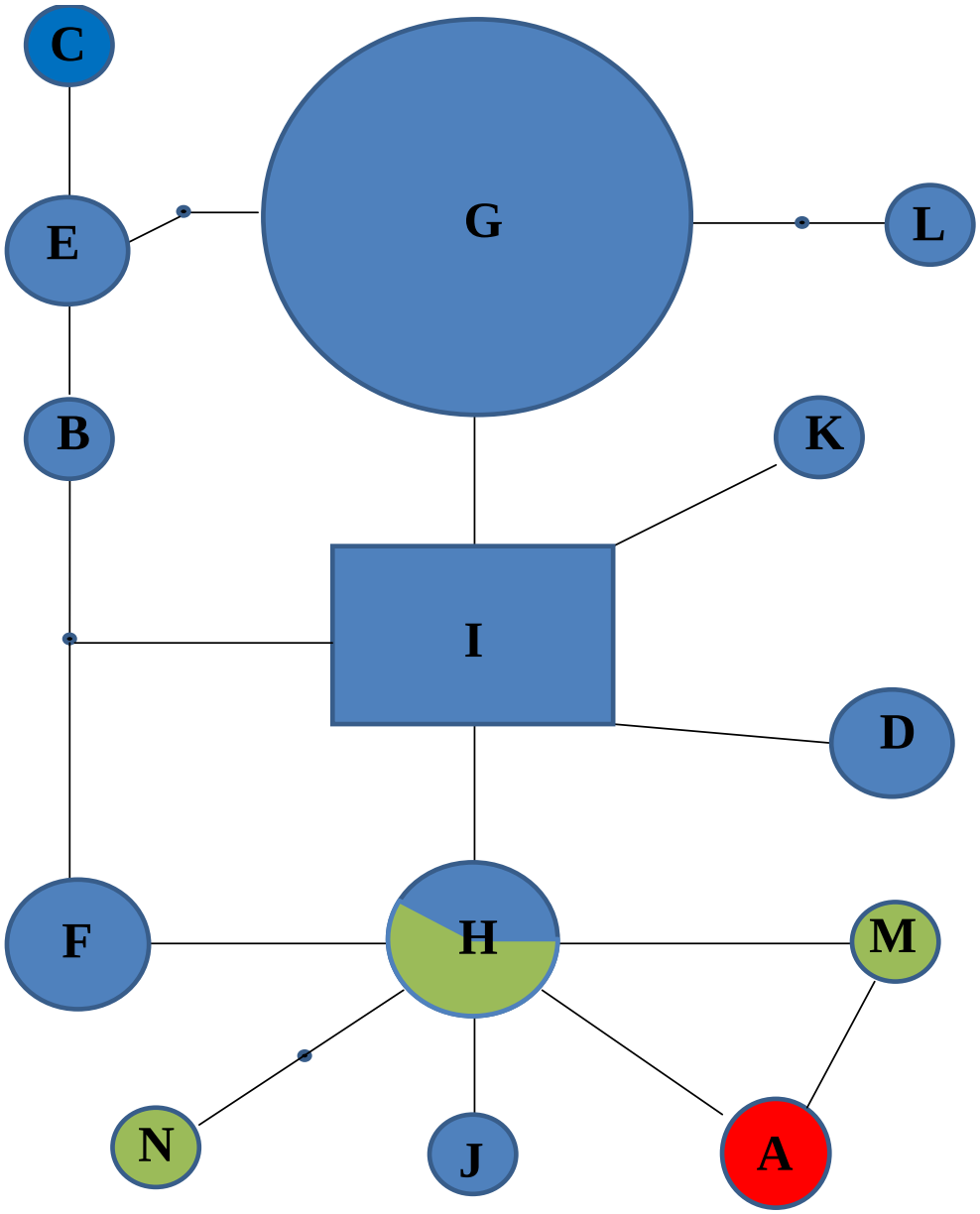


Figure 9 : Minimum Spanning Haplotype Network for Santa Ana Speckled Dace..
Each line represents one mutational step; letters represent haplotype identity; circle size is relative to number of individuals with that haplotype; small filled circles represent unsampled haplotypes; rectangle represents inferred ancestral haplotype.
(Red = San Jacinto R., Blue = Santa Ana R., Green = San Gabriel R.).

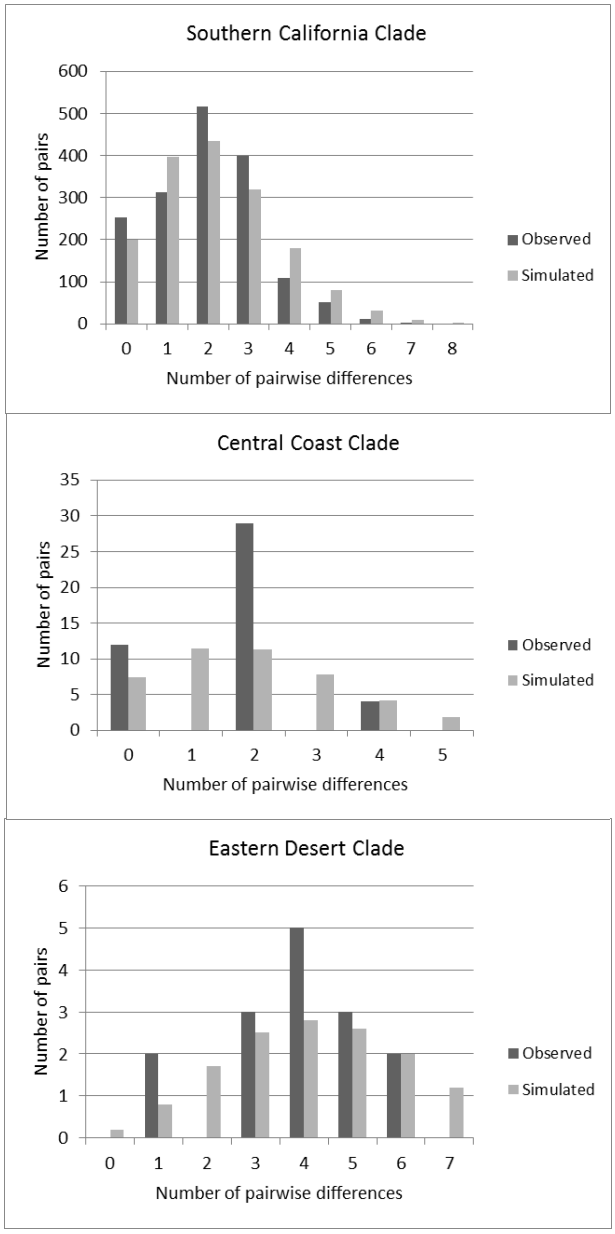


Figure 10. Mismatch distributions of pairwise nucleotide differences among *R. osculus* mtDNA haplotypes by major clades. Black bars indicate observed values, gray bars indicate simulated values fitted to the data. The Santa Ana Speckled Dace fit a model of population expansion.

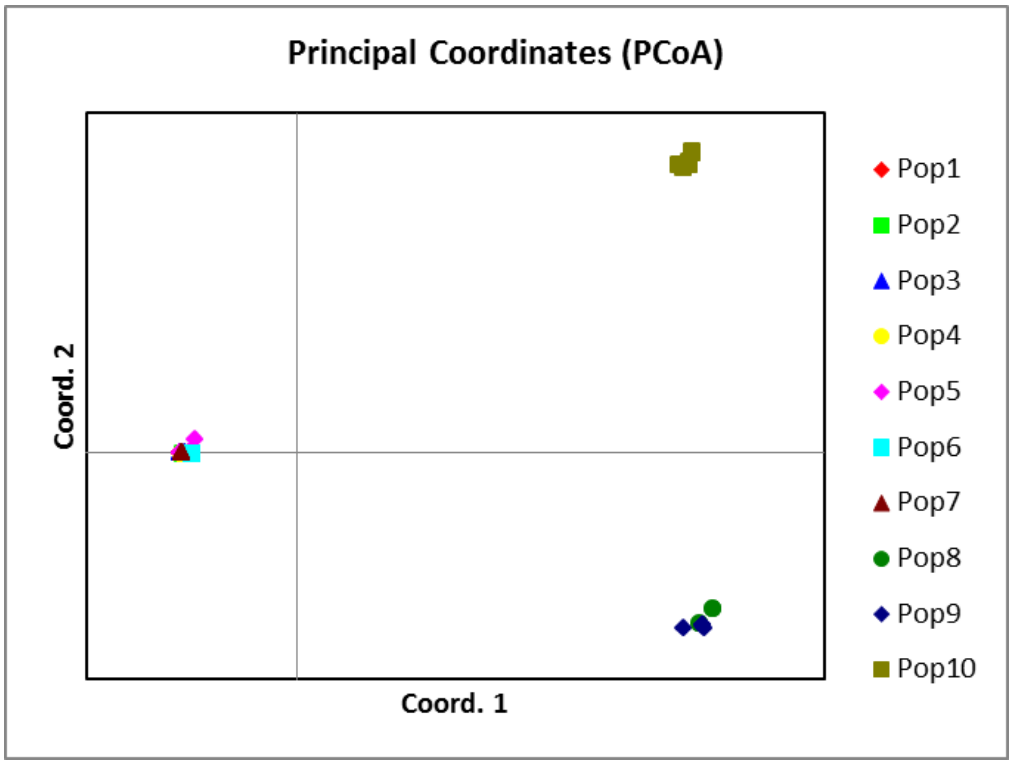


Figure 11. Principal Coordinates Analysis (PCoA) Summary. The analysis assigns each individual a location based on similarity or difference in genetic distances, geographic distances.Individuals were placed by PCoA into locations that correspond to the geographic region they occupy. Variance explained by Coord. 1= 75.1%; Variance explained by Coord. 2 = 15.9%.

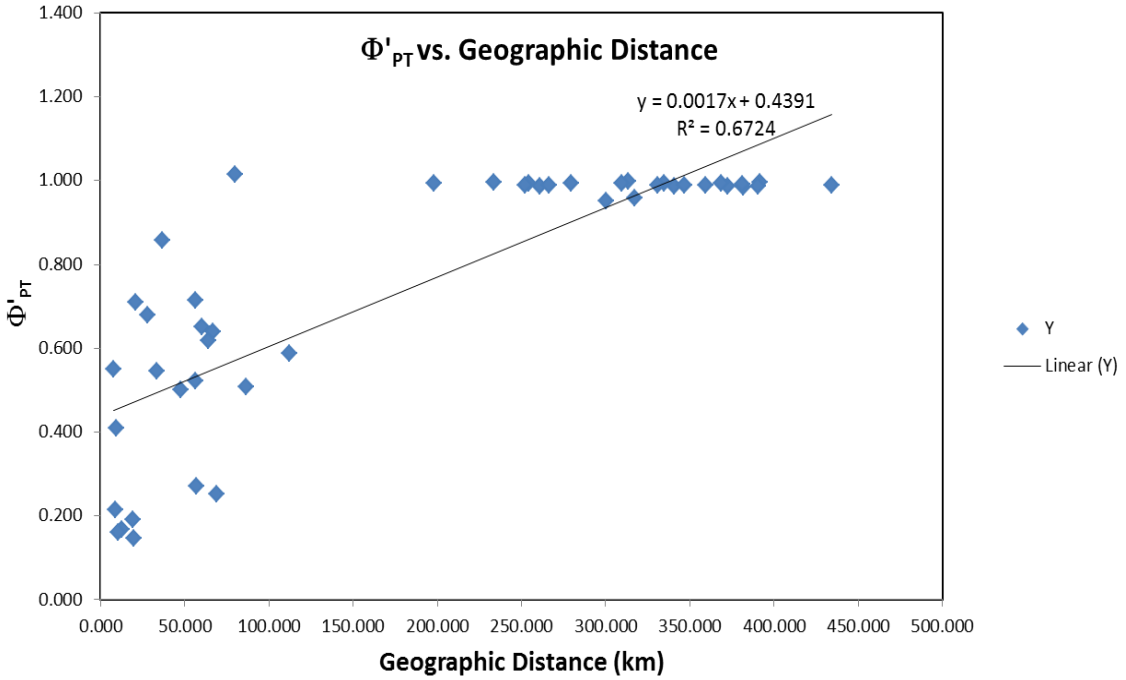


Figure 12. Mantel Test. Cumulative pairwise standardized Φ'_{PT} values for each population and geographic location for all sampling sites within California. The Mantel Test indicates strong correlation of genetic distance with geographic distance ($p < 0.001$).

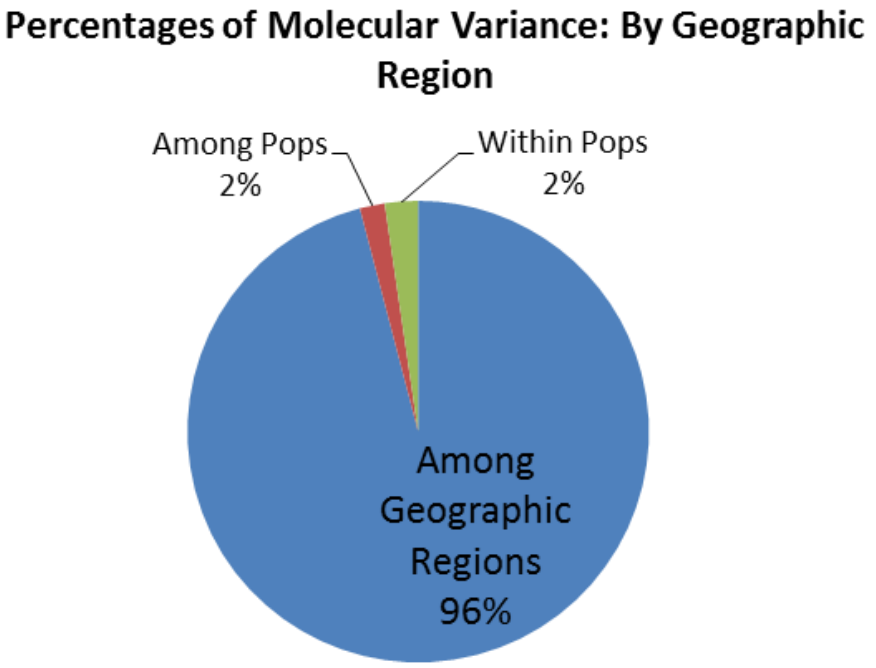


Figure 13. AMOVA Results Summary. The model showing the highest percentage of total genetic variation among speckled dace in California attributable to differences among groups was partitioned by geographic region, however all models had similar results. Significant population structure exists across geographic regions of California. (Φ_{RT} = 0.959, p = 0.001; N_m = 0.012).

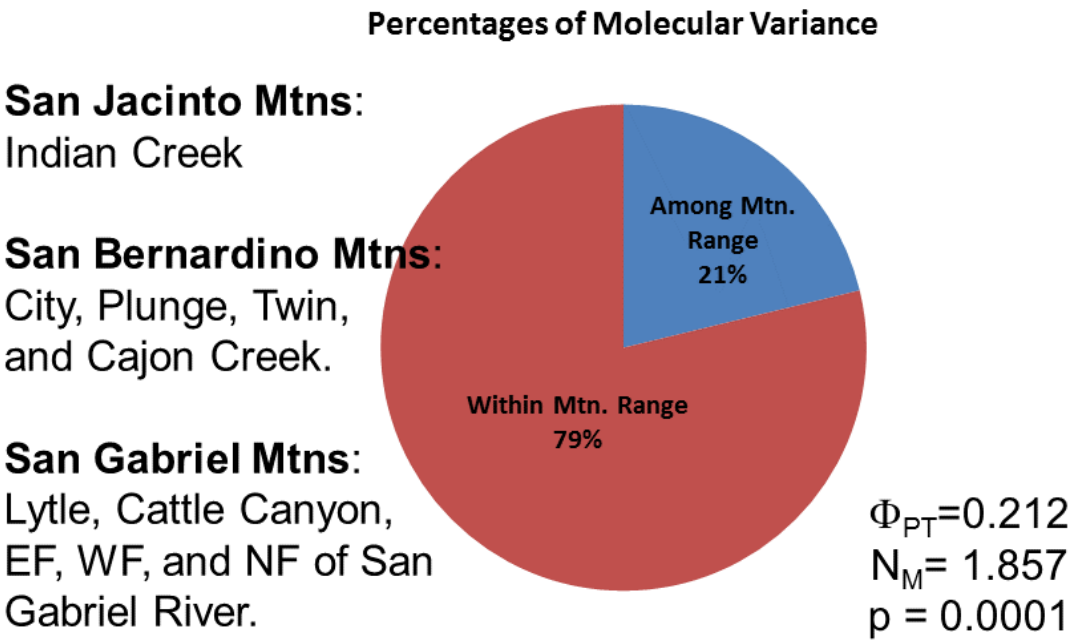


Figure 14. Model 1- Mountains. Percentage of molecular variance within the Santa Ana Speckled Dace explained by differences among mountain ranges.

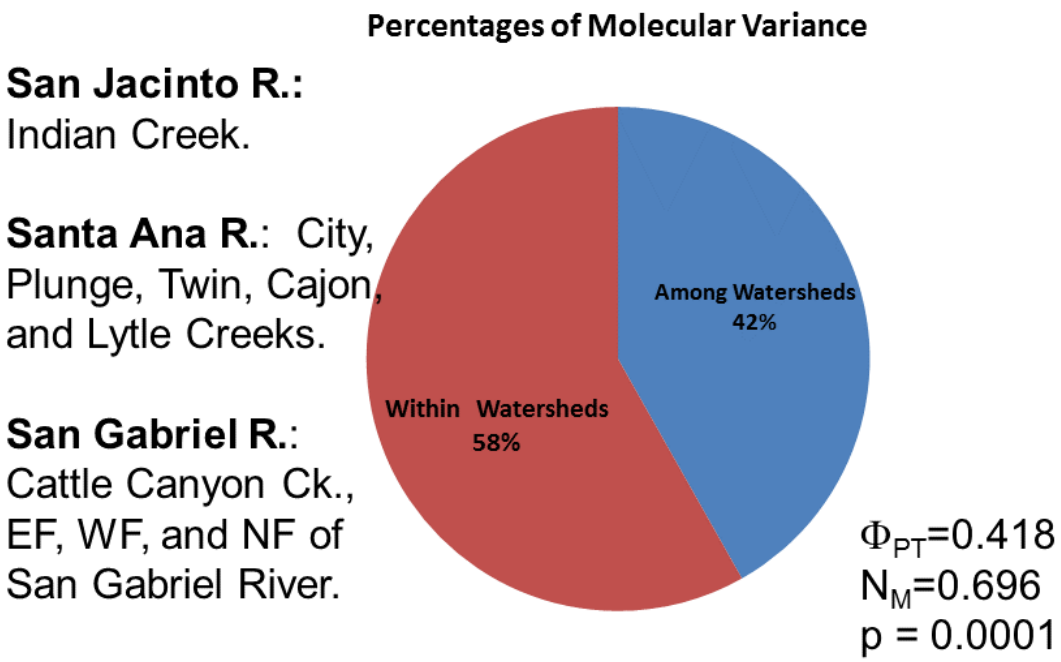


Figure 15. Model 2 - Watersheds. Percentage of molecular variance within the Santa Ana Speckled Dace explained by differences among watersheds.

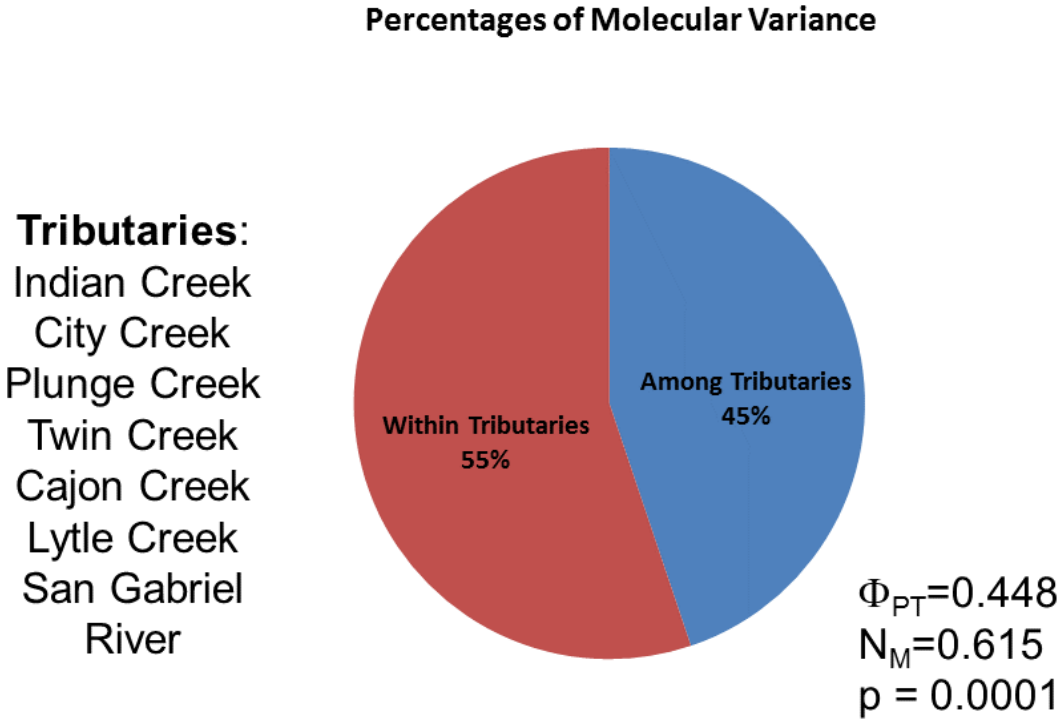


Figure 16. Model 3 - Tributary. Percentage of molecular variance within the Santa Ana Speckled Dace explained by differences among tributaries.

APPENDIX B

TABLES

Table 1. Geographic distribution of *Rhinichthys osculus* Haplotypes among sampling localities. Numbers in cells represent the number of individuals of a given haplotype (rows) at a given locality (columns).

	Locality									
Haplotype	IC	CC	PC	TWIN	CAJ	LC	SGR	SMR	SLO	OWR
A	4									
B		1								
C		1								
D		1	2							
E		1			3					
F			4		2					
G			2		7	10				
H				4			5			
I				6						
J				1						
K				1						
L					1					
M							1			
N							1			
O									2	
P									2	
Q								5		
R								1		
S										1
T										1
U										1
V										1
W										1
X										1
Total	4	4	8	12	13	10	7	6	4	6

Magenta represents the Southern California Region; green represents the Central Coast Region; blue represents the Eastern California Desert region.

Table 2. DNA Sequencing Primers for MtDNA Control Region of Speckled Dace

Locus	Name	5’ – 3’ Sequence	Source
1. D-loop	ESTFOR	catcggtcttgtaatccgaagat	Gilles et al. (2001)
2. D-loop	PHE1R	acatcttcagtgttacgctt	Gilles et al. (2001)
3. D-loop	dacedloop700R	atccgagggtggagtcttat	This study (2007)
4. D-loop	dacedloop800F	ttacatctcagagtgcaggc	This study (2007)

Table 3a. Speckled Dace Overall mtDNA Base Frequency Statistics for all Haplotypes

		Overall (1143 bp)	tRNA thr+ (48 bp)	tRNA pro- (70 bp)	mtDNA CR (986 bp)	tRNA phe+ (39 bp)
	T	0.301	0.229	0.237	0.314	0.231
Base	C	0.234	0.313	0.292	0.227	0.177
Frequencies	A	0.314	0.226	0.340	0.316	0.340
	G	0.150	0.233	0.131	0.144	0.250
Substitutions		92	1	3	85	3
Parsimony-Informative Substitutions		80	1	3	75	1

Table 3b. Speckled Dace Mean Control Region Gene Statistics for all Individuals

		Overall (986 bp)	Domain I (302 bp)	Domain II (323 bp)	Domain III (361 bp)
	T	0.314	0.319	0.335	0.289
Base	C	0.227	0.197	0.214	0.267
Frequencies	A	0.316	0.351	0.261	0.334
	G	0.144	0.133	0.189	0.110
Substitutions		85	34	19	32
Parsimony - Informative sites		75	30	17	28
	ti	59	20	13	26
	tv	27	12	7	8
	mean ti/tv	2.19	1.67	1.86	3.25

Table 4. Chi-Square Analysis of Nucleotide Substitutions by Control Region Domain

tRNAs Obs.	CR Domain I Obs.	CR Domain II Obs.	CR Domain III Obs.
7	34	19	32
tRNAs Exp.	CR Domain I Exp.	CR Domain II Exp.	CR Domain III Exp.
12.6	24.3	26	29.1

Region	Observed	Expected	Observed-Expected	(O-E)2
tRNAs	7	12.6	-5.6	2.49
CR Domain I	34	24.3	9.7	3.87
CR Domain II	19	26	-7	1.88
CR Domain III	32	29.1	2.9	0.29
			Chi-Square	8.53
			p-value	0.04

Table 5. Summary of Best-fit Evolutionary Models by mtDNA region

mtDNA Nucleotide Sequence:		Best-fit Evolutionary Model
tRNAs		Kimura 2-parameter (K2)
Control Region:		
	Overall:	Tamura 3-parameter with gamma distribution (T92 + G) model
	Domain I:	Tamura 3-parameter with gamma distribution (T92 + G) model
	Domain II:	Tamura 3-parameter with gamma distribution (T92 + G) model
	Domain III:	Tamura 3-parameter with gamma distribution (T92 +G) model
tRNAs/CR I/II/III:		Tamura 3-parameter with gamma distribution (T92 + G) model

Table 6. Results of Population statistics for Santa Ana Speckled Dace, All Populations.

		Diversity Indices		Neutrality Tests		Mismatch Distribution		
		Haplotype Diversity	ucleotide Diversity	Tajima's D	Fu's Fs	P		
Region	Number of localities	(+/- 95% CI)	(+/- 95% CI)	(P-value)	(P-value)	(SSDobs)	Raggedness	P (Ragobs)
Southern California	7	0.8475 (+/- 0.0322)	0.001761 (+/- 0.001116)	-0.8369 (0.2145)	-5.30392 (0.0125)	0.0955	0.054	0.2835
Central Coast	2	0.7333 (+/- 0.1199)	0.000974 (+/- 0.000782)	-0.8223 (0.2345)	0.16615 (0.5155)	0.018	0.918	0.005
Eastern Desert	1	1 (+/- 0.0962)	0.002637 (+/- 0.001843)	-0.1247 (0.4765)	-2.69605 (0.0300)	0.3915	0.1333	0.509

Table 7. Santa Ana Speckled Dace (*Rhinichthys osculus*) Genetic Study Sampling Localities in California. N is the sample size for each sampling location. The geographic groupings (Geographic Region, Watershed, and Mountain Ranges) for AMOVA are designated.

N	Sample Location	Lat./Long. Coordinates	Date Collected	Geographic Grouping		
				Geographic Region	Watershed	Mountain Range
13	Cajon Ck	34.25167, -117.26750	2/20/2003	Southern CA	Santa Ana	San Bernardino
10	Lytle Ck	34.23111, -117.48306	10/4/2002	Southern CA	Santa Ana	San Gabriel
12	Twin Ck	34.16694, -117.26750	4/1/2003	Southern CA	Santa Ana	San Bernardino
4	City Ck	34.18694, -117.18528	7/17-11/21/2003	Southern CA	Santa Ana	San Bernardino
3	Plunge Ck	34.11472, -117.14139	11/19/2003	Southern CA	Santa Ana	San Bernardino
5	Plunge Ck	34.11056, -117.14889	11/19/2003	Southern CA	Santa Ana	San Bernardino
4	Indian Ck	33.80644, -116.78123	11/27/2002	Southern CA	Santa Ana	San Jacinto
3	Cattle Canyon Ck	34.23472, -117.73327	8/28/2005	Southern CA	San Gabriel	San Gabriel
2	East Fork SGR	34.26007, -117.74684	8/27/2005	Southern CA	San Gabriel	San Gabriel
1	North Fork SGR	34.25258, -117.85798	8/27/2005	Southern CA	San Gabriel	San Gabriel
1	West Fork SGR	34.24066, -117.88161	8/27/2005	Southern CA	San Gabriel	San Gabriel
4	San Luis Obispo Ck	35.27679, -120.66952	5/11-6/6/2003	Central Coast CA	San Luis Obispo	Coast Range
4	Sisquoc River	34.7719, -119.94438	8/30/2005	Central Coast CA	Santa Maria	Coast Range
2	Cuyama River	35.02191, -120.22849	8/30/2005	Central Coast CA	Santa Maria	Coast Range
6	Owen's River	37.41652, -118.61874	4/5-5/16/2003	Eastern Desert CA	Owen's	Sierra Nevada

Table 8. Variable Sites within the 1143-Base Pair *Rhinichthys osculus* mtDNA Haplotypes.

Haplotype	N	tRNA thr (48bp)	tRNA pro (70bp)	Control Region Domain I (302bp)	Control Region Domain II (323bp)
				1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 4 4 4 4	4 4 4 5 5 5 5 5 5 5 5 6 6 6 7 7 7 7 7
		3	5 5 7	2 3 4 5 6 7 7 7 8 8 9 4 7 8 8 8 8 9 0 0 0 1 2 3 4 6 6 7 9 9 0 0 0 1	4 8 9 1 1 4 4 4 5 6 7 2 3 9 0 1 2 3 4
Haplotype	N	3	7 8 2	3 8 4 5 5 2 3 7 2 3 0 9 6 3 5 7 9 0 0 1 3 0 8 3 6 5 6 6 1 3 0 7 9 0	6 9 8 1 6 2 3 7 0 3 9 7 2 8 2 0 3 2 2
cr Haplotype A	4	A	G A T	G G A C C C A C T C T C A A A C A A A G C G A T C A T T A A A T A T	T T A T C G T T A C G C A G T T T A G
cr Haplotype B	1	.	. . .	. . . . C	. . . . T . . . . A . . . . .
cr Haplotype C	1	.	. . .	. . . . . . G . . . . C	. . . . T . . . . G A . . . . .
cr Haplotype D	3	.	. . .	. . . .	. . . . A . . . . .
cr Haplotype E	4	.	. . .	. . . . C	. . . . T . . . . G A . . . . .
cr Haplotype F	6	.	. . .	. . . . C	. . . .
cr Haplotype G	19	.	. . .	. . . .	. . . . G A . . . . .
cr Haplotype H	9	.	. . .	. . . .	. . . .
cr Haplotype I	6	.	. . .	. . . .	. . . . A . . . . .
cr Haplotype J	1	.	. . .	. . . .	. . . .
cr Haplotype K	1	.	. . .	. . . .	. . . . T . A . . . . .
cr Haplotype L	1	.	. . .	. . . .	. . . . G A . . . . .
cr Haplotype M	1	.	. . .	. . . .	. . . .
cr Haplotype N	1	.	. . .	. . . . G . . . . T	. . . .
cr Haplotype O	2	G	. G C	T C G T T T . A C T C . T . G . . . . T C . T C C G . . A G .	C C T C . C C . . . . . G A G . C . C
cr Haplotype P	2	G	. G C	T C G T T T . A C T C . T . G . . . . T C . T C C G . . A G .	C C T C . C C . . . . . G A G . C C C
cr Haplotype Q	5	G	. G C	T C G T T T . A C T C . T . G . . . . T C . T C C G . . A G .	C C T C . C C . . . . . G A G . C . C
cr Haplotype R	1	G	. G C	T C G T T T . A C T C . T . G . G . . . . . T C . T C C G . . A G .	C C T C . C C . . . . . G A G . C . C
cr Haplotype S	1	.	A G C	T . G T T . . A C T C . . T G T . T . . G A T . . . . G A G C	. . T C . A C . T T T . G A G . C C C
cr Haplotype T	1	.	A G C	T . G T T . . A C T C . . T G T . T G A G A T . . . . G A G C	. . T C . A C . T T T . G A G C C C C
cr Haplotype U	1	.	A G C	T . G T T . . A C T C . . T G T . T G A G A T . . . . G A G C	. . T C . A C . T T T . G A G C C C C
cr Haplotype V	1	.	A G C	T . G T T . . A C T C . . T G T . T . . G A T . . . . G A G C	. . T C . A C . T T T . G A G C C C C
cr Haplotype W	1	.	A G C	T . G T T . . A C T C . . T G T . T . . G A T . . . . G A G C	. . T C . A C C T T T . G A G . C C C
cr Haplotype X	1	.	A G C	T . G T T . . A C T C . . T G T . T . . G A T . . . . G A G C	. . T C . A C . T T T . G A G . C C C

Haplotype	N	Control Region Domain III (361 bp)	tRNA phe (39bp)
		. . . . 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	1 1 1
		7 7 7 8 8 8 8 9 9 9 9 9 9 9 9 9 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1	1 1 1
		5 7 8 4 5 6 7 3 4 4 4 5 6 7 9 9 1 1 2 3 3 3 4 5 6 6 6 6 8 9 9 0	1 2 2
	N	6 4 0 5 6 3 3 1 2 5 6 5 6 9 5 8 0 9 7 4 5 9 8 9 2 3 5 9 8 1 5 3	5 1 2
cr Haplotype A	4	A A T T C T C C A A A T G G C A G T C G A G A T A C G A T G G C	C G C
cr Haplotype B	1	. . . . A	. . .
cr Haplotype C	1	. . . . A	. . .
cr Haplotype D	3	. . . . G . . . A	. . .
cr Haplotype E	4	. . . . A	. . .
cr Haplotype F	6	. . . . A	. . .
cr Haplotype G	19	. . . . A	. . .
cr Haplotype H	9	. . . . A	. . .
cr Haplotype I	6	. . . . A	. . .
cr Haplotype J	1	. . . . A A	. . .
cr Haplotype K	1	. . . . A	. . .
cr Haplotype L	1	. . . . A . A . . . . . G	. . .
cr Haplotype M	1	. . . . T	. . .
cr Haplotype N	1	. . . . A	. . .
cr Haplotype O	2	G G C . T A T T G G G . A A T . . . T A T . . A . T A . A C . T	. . .
cr Haplotype P	2	G G C . T A T T G G G . A A T . . . T A T . . A G T A . A C . T	. . .
cr Haplotype Q	5	G G C . T A T T G G G . A A T . . . T A T . . A G T A . A C . T	. . .
cr Haplotype R	1	G G C . T A T T G G G C A A T . . . T A T . . A G T A . A C . T	. . .
cr Haplotype S	1	G . A C T A . T G . G . A A T . A C T A T A T . G T A . A C A T	. A T
cr Haplotype T	1	G . A C T A T T G . G . A A T . A C T A T A T . G T A . A C A T	. A .
cr Haplotype U	1	G . A C T A T T G . G . A A T . A C T A T A T . G T A . A C A T	T A .
cr Haplotype V	1	G . A C T A T T G . G . A A T . A C T A T A T . G T A . A C A T	. A .
cr Haplotype W	1	G . A C T A T T G . G . A A T . A C T A T A T . G T A . A C A T	. A .
cr Haplotype X	1	G . A C T A T T G . G . A A T . A C T A T A T . G T A . A C A T	. A .

R. osculus haplotype frequency and variable site data for the 1143 bp sequences. Numbers above columns indicate nucleotide position relative to 1143 bp sequence; dots indicate identity with haplotype A.

Table 9. Results of AMOVA analysis of genetic variation within and among populations of Rhinichthys osculus partitioned by geographic models based on watersheds, mountain ranges, geographic regions. All Phi statistics were found significant (P<0.01)

			Model			
	Watersheds (5 groups)		Mountains (5 groups)		Regions (3 groups)	
	% total variation	Φ-statistics	% total variation	Φ-statistics	% total variation	Φ-statistics
Among groups	94.75	Φ-RT = .95	92.91	Φ-RT = .93	96.02	Φ-RT = .96
Among populations within groups	2.20	Φ-PR = .42	3.08	Φ-PR = .43	1.74	Φ-PR = .44
Within populations	3.05	Φ-PT = .97	4.01	Φ-PT = .96	2.24	Φ-PT = .98

Table 10. AMOVA Results Summary of Pairwise Φ_{PT} Values

Pairwise Population PhiPT Values

IC	CC	PC	TWIN	CAJ	LC	SGR	SLO	SMR	OWR	
0	0.034	0.004	0.002	0.001	0.001	0.01	0.03	0.007	0.003	IC
0.704	0	0.083	0.001	0.083	0.002	0	0.03	0.007	0.007	CC
0.493	0.211	0	0.037	0.016	0.002	0.01	0	0.001	0.003	PC
0.647	0.542	0.166	0	0.001	0.001	0	0	0.001	0.001	TWIN
0.63	0.158	0.188	0.404	0	0.052	0	0	0.001	0.001	CAJ
1	0.675	0.54	0.707	0.145	0	0	0	0.001	0.001	LC
0.585	0.609	0.249	0.27	0.513	0.855	0	0.01	0.001	0.001	SGR
0.989	0.966	0.97	0.983	0.971	0.994	0.98	0	0.007	0.001	SLO
0.986	0.969	0.97	0.982	0.971	0.991	0.98	0.5	0	0.009	SMR
0.964	0.948	0.958	0.972	0.963	0.978	0.97	0.92	0.933	0	OWR

PhiPT Values below diagonal. Probability, P(rand >= data) based on 9999 permutations is shown above diagonal. Red values are not significant.

Pairwise Population N_m (Haploid) Values Based on PhiPT Values

IC	CC	PC	TWIN	CAJ	LC	SGR	SLO	SMR	OWR	
0										IC
0.211	0									CC
0.514	1.872	0								PC
0.273	0.423	2.512	0							TWIN
0.293	2.67	2.162	0.738	0						CAJ
0	0.241	0.426	0.208	2.957	0					LC
0.354	0.321	1.506	1.354	0.474	0.085	0				SGR
0.006	0.017	0.016	0.008	0.015	0.003	0.01	0			SLO
0.007	0.016	0.015	0.009	0.015	0.004	0.01	0.5	0		SMR
0.019	0.027	0.022	0.014	0.019	0.011	0.02	0.04	0.036	0	OWR

Table 11. Conserved Sequences within the *Rhinichthys osculus* 1143 bp haplotypes. Dots indicate identity with haplotype A.

Haplotype	Termination Associated Sequence (Domain I)	Conserved Sequence Block F (Domain II)	Conserved Sequence Block E (Domain II)
A	T A T G T A T T A T C A C C A T T C A T	A T G T A G T A A G A G A C C A C C	A T G A T A G A A T C A G G G A C A C
B	. . . . .	. . . . .	. . . . .
C	. . . . .	. . . . .	. . . . .
D	. . . . .	. . . . .	. . . . .
E	. . . . .	. . . . .	. . . . .
F	. . . . .	. . . . .	. . . . .
G	. . . . .	. . . . .	. . . . .
H	. . . . .	. . . . .	. . . . .
I	. . . . .	. . . . .	. . . . .
J	. . . . .	. . . . .	. . . . .
K	. . . . .	. . . . .	. . . . .
L	. . . . .	. . . . .	. . . . .
M	. . . . .	. . . . .	. . . . .
N	. . . . . G	. . . . .	. . . . .
O	. . . . .	. . . . .	. . . . .
P	. . . . .	. . . . .	. . . . .
Q	. . . . .	. . . . .	. . . . .
R	. . . . .	. . . . .	. . . . .
S	. . . . .	. . . . .	. . . . .
T	. . . . .	. . . . .	. . . . .
U	. . . . .	. . . . .	. . . . .
V	. . . . .	. . . . .	. . . . .
W	. . . . .	. . . . .	. . . . .
X	. . . . .	. . . . .	. . . . .

Haplotype	Conserved Sequence Block D (Domain II)	Conserved Sequence Block 1 (Domain III)	Conserved Sequence Block 2 (Domain III)	Conserved Sequence Block 3 (Domain III)
A	T T A T A C T G G C A T C T G A T T A A	A G G T T C A T C A T T A A A A G A C A T A	C A A A C C C C C C T A C C C C	T G T C A A A C C C C G A A A C C A
B	. . . . .	. . . . .	. . . . .	. . . . .
C	. . . . .	. . . . .	. . . . .	. . . . .
D	. . . . .	. . . . .	. . . . .	. . . . .
E	. . . . .	. . . . .	. . . . .	. . . . .
F	. . . . .	. . . . .	. . . . .	. . . . .
G	. . . . .	. . . . .	. . . . .	. . . . .
H	. . . . .	. . . . .	. . . . .	. . . . .
I	. . . . .	. . . . .	. . . . .	. . . . .
J	. . . . .	. . . . .	. . . . .	. . . . .
K	. . . . .	. . . . .	. . . . .	. . . . .
L	. . . . .	. . . . .	. . . . .	. . . . .
M	. . . . .	. . . . .	. . . . .	. . . . .
N	. . . . .	. . . . .	. . . . .	. . . . .
O	. . . . .	. . . . . G	. . . . . T	. . . . .
P	. . . . .	. . . . . G	. . . . . T	. . . . .
Q	. . . . .	. . . . . G	. . . . . T	. . . . .
R	. . . . .	. . . . . G	. . . . . T	. . . . .
S	. . . . . T	. . . . . G	. . . . .	. . . . .
T	. . . . . T	. . . . . G	. . . . . T	. . . . .
U	. . . . . T	. . . . . G	. . . . . T	. . . . .
V	. . . . . T	. . . . . G	. . . . . T	. . . . .
W	. . . . . T	. . . . . G	. . . . . T	. . . . .
X	. . . . . T	. . . . . G	. . . . . T	. . . . .

APPENDIX C

INPUT FILES

MR BAYES INPUT FILE

#NEXUS

Begin data;
Dimensions ntax=30 nchar=1009;
Format datatype= DNA gap=-;
Matrix
cr_Haplotype_A
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCTCTCAAGAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC
cr_Haplotype_B
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC
cr_Haplotype_C
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACGTATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC
cr_Haplotype_D
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATGCATGTCATATCGACCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC
cr_Haplotype_E
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA

TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC

cr_Haplotype_F

ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC

cr_Haplotype_G

ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC

cr_Haplotype_H

ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC

cr_Haplotype_I

ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC

cr_Haplotype_J
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTATCTCAAAAATAGCCCAAAAAATTCTACTAAATTTA
CTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC
cr_Haplotype_K
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCTGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC
cr_Haplotype_L
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCTCTCAAAAATAACCCAAAAAATTCTACTAAATTTA
CTGGTAGATTTCTCAATGCTAAAAATTCGAACGTATTTAACC
cr_Haplotype_M
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAACTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCTCTCAATAATAGCCCAAAAAATTCTACTAAATTTA
CTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC
cr_Haplotype_N
ATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATCACATGCGTACAGTACACATATCTGGTTCAACAC
ATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCAGCATTCATTTATCTTAACCTA
AAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAAATTATTATAACCTGGGAATTA
TATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAAGGTTTATCATAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTAGCACACT
GTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAAGT
AGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT

ACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAACTAACAAGTTGAGATATCCGTTAGCCATCCGCG
TTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCTCTCAAAAATAGCCCAAAAAATTCTACTAAATTT
ACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACC

cr_Haplotype_O

ATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTACATGCGTATAGTACATATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAATTTATCATAACCTGGGAACTAT
ATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACACCTAAGGTTTAACGTAACCATATTAATGTAGTA
AGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATACGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATC
TGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTT
GGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTCAAGGTGGTACTCTTCCTTGCACGAATTAAAC
TAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAAC
TTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTT
CTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGGGCTAACAAGTTGAGATATCCATTAGCCATCCG
CATTATATATATATATATATACATGTCATATCGACCCCCCTAAAAATTTTCTCAAATATAGCCCAAAAAATTCTACTAAAA
TTATTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATC

cr_Haplotype_P

ATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTACATGCGTATAGTACATATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAATTTATCATAACCTGGGAACTAT
ATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACACCTAAGGTTTAACGTAACCATATTAATGTAGTA
AGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATACGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATC
TGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTT
GGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTCAAGGTGGTACTCTTCCTTGCCCGAATTAAA
CTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAA
CTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGT
TCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGGGCTAACAAGTTGAGATATCCATTAGCCATCC
GCATTATATATATATATATACATGTCATATCGACCCCCCTAAAAATTTTCTCAAATATAGCCCAAAAAATTCTACTAAA
ATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATC

cr_Haplotype_Q

ATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTACATGCGTATAGTACATATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAATTTATCATAACCTGGGAACTAT
ATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACACCTAAGGTTTAACGTAACCATATTAATGTAGTA
AGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATACGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATC
TGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTT
GGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTCAAGGTGGTACTCTTCCTTGCACGAATTAAAC
TAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAAC
TTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTT
CTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGGGCTAACAAGTTGAGATATCCATTAGCCATCCG
CATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAATTTTCTCAAATATAGCCCAAAAAATTCTACTAAA
ATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATC

cr_Haplotype_R

ATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTACATGCGTATAGTACATATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCATGTACTAACGTCCGAGACGTACATAGACCAAATCGTTAAAACTCACAAATAATTTATCATAACCTGGGAACTAT
ATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACACCTAAGGTTTAACGTAACCATATTAATGTAGTA
AGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATACGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATC
TGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTT
GGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTCAAGGTGGTACTCTTCCTTGCACGAATTAAAC
TAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAAC
TTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTT
CTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGGGCTAACAAGCTGAGATATCCATTAGCCATCCG
CATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAATTTTCTCAAATATAGCCCAAAAAATTCTACTAAA
ATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATC

cr_Haplotype_S

ATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATTACATGCGTATAGTACACATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAGGGTTTAACGCAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACT

GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAAGACTCCACCTTCGGATAATTATACTTGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTCAAGGTGGTACTCTTCCTTGCCCGAATTAAACT
AGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACT
TACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTC
TCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAGCTAACAAGTTGAGATATCCATTAGCCATCCGC
ATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATTTTCTCAAATATAACCCAAAAATTTCTACTAAATT
TGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATC

cr_Haplotype_T

ATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATTACATGCGTATAGTACACATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTATCGTTCATGACGTACATGAAGCAAATCATTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAGGGTTTAACGCAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAAGACTCCACCTTCGGATAATTATACTTGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCCCAAGGTGGTACTCTTCCTTGCCCGAATTAAACT
AGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACT
TACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTTC
TCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAGCTAACAAGTTGAGATATCCATTAGCCATCCGC
ATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATTTTCTCAAATATAACCCAAAAATTTCTACTAAATT
TGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATC

cr_Haplotype_U

ATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATTACATGCGTATAGTACACATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTATCGTTCATGACGTACATGAAGCAAATCATTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAGGGTTTAACGCAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAAGACTCCACCTTCGGATAATTATACTTGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCCCAAGGTGGTACTCTTCCTTGCCCGAATTAAACT
AGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACT
TACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTTC
TCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAGCTAACAAGTTGAGATATCCATTAGCCATCCGC
ATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATTTTCTCAAATATAACCCAAAAATTTCTACTAAATT
TGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATC

cr_Haplotype_V

ATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATTACATGCGTATAGTACACATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAGGGTTTAACGCAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAAGACTCCACCTTCGGATAATTATACTTGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCCCAAGGTGGTACTCTTCCTTGCCCGAATTAAACT
AGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACT
TACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTTC
TCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAGCTAACAAGTTGAGATATCCATTAGCCATCCGC
ATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATTTTCTCAAATATAACCCAAAAATTTCTACTAAAT
TTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATC

cr_Haplotype_W

ATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATTACATGCGTATAGTACACATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAGGGTTTAACGCAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATACTTTTAAGACTCCACCTTCGGATAATTATACTTGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTCAAGGTGGTACTCTTCCTTGCCCGAATTAAACT
AGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACT
TACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTTC
TCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAGCTAACAAGTTGAGATATCCATTAGCCATCCGC
ATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATTTTCTCAAATATAACCCAAAAATTTCTACTAAATT
GTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATC

cr_Haplotype_X

ATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATTACATGCGTATAGTACACATATATGGTCTAACACA

CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACACCTAGGGTTTAACGCAACCATATTAATGTAGTA
AGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAAGACTCCACCTTCGGATAATTATACTTGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTCAAGGTGGTACTCTTCCTTGCCCGAATTAAACT
AGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTATCAAGTGCATAACATATACATCCTTTCTTCAACT
TACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTTC
TCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAGCTAACAAGTTGAGATATCCATTAGCCATCCGC
ATTATATATATATATATATACATGTCATATCAACCCCCCAAAAATTTTCTCAAATATAACCCAAAAATTTCTACTAAATTT
GTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATC

Rhinichthys_osculus_gi224813784

ATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATTACATGCGTATAGTACATATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTAACATCCAAGACGTACATGGAGCAAATCGTTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATACGGCACACAACACTTTCCTTGAAATACACACCTAAGGTTTAACGTAACCATATTAATGTAGTA
AGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACT
GTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACCTCTTTTTAAGACTCCACCCTAGGATAATTATACTGGCATC
TGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTTT
GGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAATAACATCCCAAGGTGGTACTCTTCCTTGCACGAATTAAC
CTAGGTTCATCATTGAAAGACATAACTTAAGAATTACATTTTATTCTATCAAGTGCATACACATCTATCCTTTCTTCAAC
TTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTT
CTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGGGCTAACAAGTTGAGATATCCATTAGCCATCCG
CATTATATATATATATATATACATGTCATATCAACCCCCCCCCTAAAAAATTGGCTCAAAAATAGCCCAAAAAATTCTACTA
AATTATTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATC

Rhinichthys_falcatus_gi224813783

ATCTTCTGATAGTAACCTATATGGACCGGTACCGCGTGTAGCATTACATGCGTATAGTACATATATATGGTCTAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATTTGGCACACAACACTTTCCTTGAAATAAACACCTAAGGTTTAACGTAACCATATTAATGTAGTAA
GAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACTG
TGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGGTTCTCTTTTTTG
GTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAACT
AGGTTCATCATTGAAAGACATAACTTAAGAATTACATTTTATTCTATCAAGTGCATAACATATACATCCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAACCCCCCTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCTAAACCAAGGAAGGTTCGAGAACGTGCGGGCTAACAAGTTGAGATATCCATTAGCCATCCGCA
TTATATATATATATATATATACATGTCATATCAACCCCCCCCAAAATTTTCTCAAATATAGCCCAAAAAATTCTACTAAATTT
GTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATT

Rhinichthys_sp._JM-2008_gi193203748

ATCTTCTGATAGTAACCTATATGGTCCGGTGCCATGTATAGCATTACATGTGCACAGTACACATATATGGTCTAACACA
CACATGATATTATTCGTAATATTGTGTGTGTGTGTGTTAGGGCATATATATGTATTATCACCATTCATTTATCTTAACCT
AAAAGCAAGTACTAACATCTAAGACGTACATAGACCATATCGTTAAAACTCACAAATAATTTATTATAACCTGAGAACTA
TAATATTCCCCTAGATATGGCACACAACACTTTCCTTGAAATACACACCTAAGGTTTAACGTAACCATATTAATGTAGT
AAGAGACCACCAACCGGTCCATATAAGGCATACTATTCATGATAGAATCAGGGACACAATATGTAGATAAGGTATATT
GTGAACTATTCCTTGTATCTGGTTTCTATCTCAGGCACAGAATTGTGAAGTTCCACCCTACAGTATCGCACTTGCATCC
GGTTACTGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGTTTTTTTTTTAGG
TTTCCTTTCACTTTACATCTCAAAGTGCAGGCACAAGTAACATCTCAGGGTGGTACATTTCCTTGCACGAATTAACGTA
GGTTCATCATTGAAAGACATAACTTAAGAATTACATTATACTCTATCAAGTGCATAACATATTCATTCTTTCTTCAACTTA
CCCTGATATATACGCCCCCTCTTTTGGCTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTC
CTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTCCAAGCTAACAAGTTGAGATATCCATTAGCCATCCGCGT
TATATATATATATATATATATATACATGCCATATCCCCCGAAAAATTTCCCAAATATAGCCCAGAAAATTCTACTAAATTT
ATTGGGAAATTTCTCAATGCTAAAAATTCCAACATTATTTGGGC

Rhinichthys_evermanni_gi193081196

ATCTTCTGATAGTAACCTATATGGTCCGGTGCAGTGTATAGCATTACATGTGCACAGTACACATATATGGTCCGACAC
ACACATGATATTATTCGTAATATTGTGTGTGTGTGTGTTAGGGCATATATATGTATTATCACCATTCATTTATCTTAACC
TAAAAGCAAGTACTAACATCTAAGACGTACATAGACCATATCGTTAAAACTCACAAATAATTTATTATAACCTGGGAAG
TATAATATTCCCCTAGATATGGCACACAACACTTTCCTTGAAATACACACCTAAGGTTTAACATAACCATATTAATGTAG
TAAGAGACCACCAACAGGTCCATATAAGGCATACTATTCATGATAGAATCAGGGACACAATATGTAGATAAGGTATAT
TGTGAACTATTCCTTGTATCTGGTTTCTATCTCAGGCACAGAATTGTGAAGTTCCACCCTACAGTATCGGACTGGCAT
CCGGTTACTGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGTTTTTTTTTTA
GGTTTCCTTTCACTTTACATCTCAAAGTGCAGGCACAAGTAACATCTCAAGGTGGTACATTTCCTTGCACGAATTAACG
TAGGTTCATCATTGAAAGACATAACTTAAGAATTACATTATACTCTATCAAGTGCATAACATATTCATTCTTTCTTCAACT
TACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTTC
TCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTCCAAGCTAACAAGTTGAGATATCCATTAGCCATCCGC

GTTATATATATATATATATATATACATGCCATACCCCCCAAAAAATTTCCCAAATATAGCCCAGAAAATTCTACTAAA
TTTATTGGGAAATTTCTCAAGGCGAAAAATTCCAACATTATTTGGGC
Rhinichthys_umatilla_gi225194554
ATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATTACATGCGTATAGTACATATATATGGTCCAACACA
CACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATATATGTATTATCACCATTCATTTATCTTAACCTAA
AAGCAAGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAACTCACAAATAATTTATTATAACCTGGGAACTAT
ATTATTCCCCTAGATTTGGCACACAACACTTTCCTTGAAATAAACACCTAAGGTTTAACGTAACCATATTAATGTAGTAA
GAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAATCAGGGACACAATATGTGGGGGTTGCACACTG
TGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTATAAGACTCCACCCTCGGATAATTATACTGGCATCT
GATTAATGGTGTAATTACATACTCTCCTCGTTACCCAACATGCCGGGCGTTCTTTTATATGCATAGGGGTTCTCTTTTT
TGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAATAACATCTCAAGGTGGTACTTTTCCTTGCACGAATTAAA
CTAGGTTCATCATTGAAAAAGACATAACTTAAGAATTACATTTTATTCTATCAAGTGCATAACATATACATCCTTTCTTC
AACTTACCCTGATATATACGCCCCCCTTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCT
GTTCTCCTTGTCAAACCCCTAAACCAAGGAAGGTTCGAGAACGTGCGGGCTAACAAGTTGAGATATCCATTAGCCAT
CCGCATTATATATATATATATATACATGTCATATCAACCCCCCCCAAAATTTTCTCAAATATAGCCCAAAAAATTCTACT
AAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATC
Rhinichthys_cataractae_gi225194553
ATCTTCTGATAGTAACCTATATGGTCCGGTGCCGTGTATAGCATTACATGTGCACAGTACATATATATGGTCTAACACA
CACATGATATTATTCGTAATATTGTGTGTGTGTGTTAGGGCATATATATGTATTATCACCATTCATTTATCTTAACCT
AAAAGCAAGTACTAACATCTAAGACGTACATAGACCAAATCATTAAAACTCACAAATAATTTATTATAACCTGAGAACTA
TAATATTCCCCTAGATATGGCACACAACACTTTCCTCGAAATACACACCTAAGGTTTAACATAACCATATTAATGTAGT
AAGAGACCACCAACCGGTCCATATAAGGCATACTATTCATGATAGAATCAGGGACACAATGTGTAAATAAGGTATATT
GTGAACTATTCCTTGTATCTGGTTTCTATCTCAGGCACAGAATTATGAAGTTCCACCCTACAGTGTCGCACTGGCATC
CGGTTACTGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGCATTCTTTTATATGCATAGGGTTTTTTTTTTAG
GTTTCCTTTCACTTTACATCTCAAAGTGCAGGCACAAGTAACATCTCAAGGTGGTACATTTCCTTGCACGAATTAACGT
AGGTTCATCATTGAAAGACATAACTTAAGAATTACATTATACTCTATCAAGTGCATAACATATTCATTCTTTCTTCAACTT
ACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAACCCCCTTACCCCCTACGCTCAGCAAATCCTGTTCT
CCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTCCAAGCTAACAAGTTGAGATATCCATTAGCCATCCGCG
TTATATATATATATATATATATACATGCCATATCAACCCCCCAAAAAATTTCCCAAATATAGCCCAGAAAATTCTACTAA
ATTTATTGGTAAATTTCTCAATGCTAAAAATTCCAACATTATTTGGGC
;
end;

begin trees;
[This tree is a ML tree from the same alignment above]

Treedacedloophaplotypes=(((((((((cr_Haplotype_A:0.00068061,cr_Haplotype_M:0.00068061):0.00068061,cr_Haplotype_
H:0.00000000):0.00000000,cr_Haplotype_N:0.00204360):0.00000000,cr_Haplotype_J:0.00102110):0.00000000,(cr_Hapl
otype_F:0.00000000,(cr_Haplotype_B:0.00000000,(cr_Haplotype_C:0.00102110,cr_Haplotype_E:0.00000000):0.001021
10):0.00204360):0.00102110):0.00102110,(cr_Haplotype_D:0.00102110,(cr_Haplotype_I:0.00000000,cr_Haplotype_K:0.
00102110):0.00000000):0.00000000):0.00102110,(cr_Haplotype_G:0.00000000,cr_Haplotype_L:0.00204360):0.0000000
0):0.03304909,(((((cr_Haplotype_Q:0.00000000,cr_Haplotype_R:0.00204360):0.00000000,cr_Haplotype_P:0.00102110):
0.00000000,cr_Haplotype_O:0.00102110):0.01369716,(Rhinichthys_falcatus_gi224813783:0.00351571,Rhinichthys_uma
tilla_gi225194554:0.00674265):0.00234982):0.00139294,(Rhinichthys_osculus_gi224813784:0.01798065,(cr_Haplotype_
S:0.00000000,((cr_Haplotype_W:0.00102110,cr_Haplotype_X:0.00000000):0.00000000,(cr_Haplotype_V:0.00000000,(cr
_Haplotype_T:0.00000000,cr_Haplotype_U:0.00000000):0.00204360):0.00102110):0.00102110):0.01733412):0.0037825
0):0.00947650):0.01914255,(Rhinichthys_cataractae_gi225194553:0.00555961,(Rhinichthys_sp._JM-
2008_gi193203748:0.00516553,Rhinichthys_evermanni_gi193081196:0.00922646):0.00372959):0.04638470);

begin mrbayes;
	log start filename = dacedloophaplotypes6.log replace;
	set autoclose = no nowarn=no;
	outgroup Rhinichthys_evermanni_gi193081196;
	charset molecule = 1-1009;
	charset tRNA = 1-8;
	charset CR_1 = 9-311;
	charset CR_2 = 312-636;
	charset CR_3 = 637-1009;
	partition dacecr= 4: tRNA, CR_1, CR_2, CR_3;
	partition all= 1: molecule;
	set partition=dacecr;
	lset nst=2;
	prset ratepr=variable;
		[For the JC model add to the prset command the option statefreqpr = fixed(equal]
	unlink revmat=(all) shape=(all) pinvar=(all)statefreq=(all) tratio= (all);
	showmodel;

```
mcmc ngen=30000000 printfreq=1000 samplefreq=100 nchains=4 temp=0.2 checkfreq = 50000 diagnfreq =
1000 stopval = 0.01 stoprule = yes;
sumt relburnin = yes burninfrac = 0.25 contype = halfcompat;
sump relburnin = yes burninfrac =0.25;
log stop;
end;
```

GENALEX FASTA INPUT FILE

```
>cr_IC1
```

CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_IC2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_IC3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_IC4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC

ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CC2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACGTATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CC6
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATGCATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CC9
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA

ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PC4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PC7
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATGCATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PC10
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PC11
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA

TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PC12
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PC13
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PC14
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATGCATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC

ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTAT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN5
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCTGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC

AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN6
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN7
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN8
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN9
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC

TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN10
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN11
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_TWIN12
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT

CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ5
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ6
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAACCCAAAAAATTCTACTAAATTTACTGGTAGATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT

>cr_CAJ7
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ8
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ9
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ10
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ11

CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ13
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ14
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CAJ15
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC

ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK5
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA

ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK6
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK7
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK8
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK9
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA

TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_LCK10
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CATC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CATC2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CATC3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC

ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAATAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_EFSGR1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_EFSGR3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCAGCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAATTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_NFSGR2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_WFSGR1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC

AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_STC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCTAAAAATT
TTCTCAAATATAGCCCAAAAAATTCTACTAAAATTATTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCGCT
AGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_STC2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCTAAAAATT
TTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCGC
TAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_BC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCTAAAAATT
TTCTCAAATATAGCCCAAAAAATTCTACTAAAATTATTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCGCT
AGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_SLO1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG

GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCTAAAAATT
TTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCGC
TAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_MNCR1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAAT
TTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCG
CTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_MNCR2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAAT
TTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCG
CTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CUYR1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCGAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGCTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAA
TTTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCG
CTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_CUYR2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAATTT

TCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCGCT
AGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_DBCR1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAAT
TTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCG
CTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_DBCR2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAAT
TTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCG
CTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
>cr_PNC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTC
AAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTAT
CAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACC
CCCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAG
CTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATACATGTCATATCAACCCCCCCAAAAATTT
TCTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCTA
GCGTAGCTTAATATAAAGCGTAACACTGAAGATGT
>cr_PNC2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATGAAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCC
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGA
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATT
TTCTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCT
AGCGTAGCTTAATACAAAGCGTAACACTGAAGATGT

>cr_PNC3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATGAAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCC
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGA
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATT
TTCTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCT
AGCGTAGTTTAATACAAAGCGTAACACTGAAGATGT
>cr_MM1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCC
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGA
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCCCAAAAAT
TTTCTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGC
TAGCGTAGCTTAATACAAAGCGTAACACTGAAGATGT
>cr_MM2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATACTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTC
AAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTAT
CAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACC
CCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAG
CTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCAAAAATTTT
CTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCTAG
CGTAGCTTAATACAAAGCGTAACACTGAAGATGT
>cr_MM3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTC
AAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTAT
CAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACC
CCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAG
CTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCAAAAATTTT
CTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCTAG
CGTAGCTTAATACAAAGCGTAACACTGAAGATGT

ARLEQUIN INPUT FILE

```
#================================
#Project file created by Arlequin
```

```
#================================

#    (symbol # is used to comment what follows on a line)

[Profile]
    Title="Dace MtDNA Control Region Analysis"
    NbSamples=10
    DataType=DNA  # - {DNA, RFLP, MICROSAT, STANDARD, FREQUENCY}
    GenotypicData=0  # - {0, 1}
    GameticPhase=0  # - {0, 1}
    LocusSeparator=NONE  # - {TAB, WHITESPACE, NONE}
    RecessiveData=0  # - {0, 1}
    MissingData='?'  # A single character specifying missing data
#    Some advanced settings the experienced user can uncomment
#    Frequency= ABS            # - {ABS, REL}
#    FrequencyThreshold= 1.0e-5 # - (Any real number, usually between 1.0e-7 and 1.e-3)
#    EpsilonValue= 1.0e-7       # - (Any real number, usually between 1.0e-12 and 1.0e-5)

[Data]

[[HaplotypeDefinition]]
HaplListName="Dace MtDNA Control Region Haplotype List"
HaplList={
A
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
B
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
C
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACGTATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
```

AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
D
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATGCATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
E
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
F
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
G
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC

TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
H
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
I
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
J
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTAT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
K
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCTGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT

CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
L
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAACCCAAAAAATTCTACTAAATTTACTGGTAGATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
M
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAATAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
N
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCAGCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAATTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
O
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGA-
CCCCCCTAAAAATTTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTATTAGTAAATTTCTCAATGCTAAAAAATCCA
ACGTATTTAATCGCTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT

P
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGA-
CCCCCCTAAAAATTTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCA
ACGTATTTAATCGCTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
Q
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCAAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAAT
TTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCG
CTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
R
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAGACCCCCTCCTAGCGCCCAGAAAAGGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCCGTACCGCGTGTAGCATTA
CATGCGTATAGTACATATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATATA
TATGTATTATCACCATTCATTTATCTTAACCTAAAAGCATGTACTAACGTCCGAGACGTACATAGACCAAATCGTTAAAA
CTCACAAATAATTTATCATAACCTGGGAACTATATTATTCCCCTAGATTCGGCACACAACACTTTCCTTGAAATGAACA
CCTAAGGTTTAACGTAACCATATTAATGTAGTAAGAGACCACCAACCGGTCCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATACGTGGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTCCATATTTATA
AGACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGG
CGTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCT
CAAGGTGGTACTCTTCCTTGCACGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAGTTACACTTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGG
GCTAACAAGCTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCGACCCCCCCTAAAAA
TTTTCTCAAATATAGCCCAAAAAATTCTACTAAAATTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACGTATTTAATCG
CTAGCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
S
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTC
AAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTAT
CAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACC
CCCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAG
CTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATTT
TCTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCTA
GCGTAGCTTAATATAAAGCGTAACACTGAAGATGT
T
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC

CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATGAAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCC
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGA
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATT
TTCTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCT
AGCGTAGCTTAATACAAAGCGTAACACTGAAGATGT
U
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATGAAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCC
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGA
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCCAAAAATT
TTCTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCT
AGCGTAGTTTAATACAAAGCGTAACACTGAAGATGT
V
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCC
CAAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTA
TCAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAAC
CCCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGA
GCTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCCCAAAAAT
TTTCTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGC
TAGCGTAGCTTAATACAAAGCGTAACACTGAAGATGT
W
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATACTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTC
AAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTAT
CAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACC
CCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAG
CTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCAAAAATTTT
CTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCTAG
CGTAGCTTAATACAAAGCGTAACACTGAAGATGT
X
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAAGGAGATTCTAACTCCCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGTAACCTATATGGTCCGGTACCGCGTGTAGCATT
ACATGCGTATAGTACACATATATGGTCTAACACACACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT

```
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTATCGTTCATGACGTACATAGAGCAAATCATTAAA
ACTCACAAATAATTTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAGGGTTTAACGCAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTTGCACACTGTGAACTATTCCTTGCATCTGGTTCCTATTTCAGGTACATATTTTTAA
GACTCCACCTTCGGATAATTATACTTGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAAGTAACATCTC
AAGGTGGTACTCTTCCTTGCCCGAATTAAACTAGGTTCATCATTGAAAGACATAACTTAAGAATTACAATTTACTCTAT
CAAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCCCTTTTGGTTTTCACGCGACAAACC
CCCTTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGTTCGAGAACGTGCGAG
CTAACAAGTTGAGATATCCATTAGCCATCCGCATTATATATATATATATATACATGTCATATCAACCCCCCAAAAATTTT
CTCAAATATAACCCAAAAATTTCTACTAAATTTGTTAGTAAATTTCTCAATGCTAAAAAATCCAACATATTTAATCGCTAG
CGTAGCTTAATACAAAGCGTAACACTGAAGATGT

}

[[Samples]]

    SampleName="IC"
    SampleSize=4  #Fictive number, but must match the sume of haplotype frequencies given below
              SampleData= {
      #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
            A 4
    }

    SampleName="CC"
    SampleSize=4  #Fictive number, but must match the sume of haplotype frequencies given below
              SampleData= {
      #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
            B 1
            C 1
            D 1
            E 1
    }

    SampleName="PC"
    SampleSize=8  #Fictive number, but must match the sume of haplotype frequencies given below
              SampleData= {
      #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
            D 2
            F 4
            G 2
    }

    SampleName="TWIN"
    SampleSize=12  #Fictive number, but must match the sume of haplotype frequencies given below
              SampleData= {
      #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
            H 4
            I 6
            J 1
            K 1
    }

    SampleName="CAJ"
    SampleSize=13  #Fictive number, but must match the sume of haplotype frequencies given below
              SampleData= {
      #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
      E 3
            F 2
            G 7
            L 1
    }

    SampleName="LCK"
    SampleSize=10  #Fictive number, but must match the sume of haplotype frequencies given below
              SampleData= {
```

```
  #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
        G 10
}

SampleName="SGR"
SampleSize=7  #Fictive number, but must match the sume of haplotype frequencies given below
          SampleData= {
  #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
        H 5
        M 1
        N 1
}

SampleName="SLO"
SampleSize=4  #Fictive number, but must match the sume of haplotype frequencies given below
          SampleData= {
  #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
        O 2
        P 2
}

SampleName="SMR"
SampleSize=6  #Fictive number, but must match the sume of haplotype frequencies given below
          SampleData= {
  #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
        Q 5
        R 1
}

SampleName="OWR"
SampleSize=6  #Fictive number, but must match the sume of haplotype frequencies given below
          SampleData= {
  #Example of a sample consisiting of haplotypic data (2 haplotypes, 2 loci):
        S 1
        T 1
        U 1
        V 1
        W 1
        X 1
}


        [[Structure]]

                StructureName="New Edited Structure"
                NbGroups=3

                Group={
                        "IC"
                        "CC"
                        "PC"
                        "TWIN"
                        "CAJ"
                        "LCK"
                        "SGR"
                }

                Group={
                        "SLO"
                        "SMR"
                }

                Group={
                        "OWR"
                }
```

TCS INPUT FILE

#NEXUS

Begin data;
Dimensions ntax=58 nchar=1143;
Format datatype=nucleotide gap=- missing=? matchchar=.;
Matrix
IC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
IC2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
IC3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
IC4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC

TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCCTAAAAATTCT
CTCAAGAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CC2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACGTATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CC6
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATGCATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CC9
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT

CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
PC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
PC4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
PC7
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
PC10
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT

PC11
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
PC12
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
PC13
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
PC14
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATGCATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC

CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTAT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN5
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT

ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCTGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN6
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN7
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN8
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN9
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA

CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN10
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN11
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
TWIN12
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA

GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ5
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ6
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC

AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAACCCAAAAAATTCTACTAAATTTACTGGTAGATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ7
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ8
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ9
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ10
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTTCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC

CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ11
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAACCA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ13
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ14
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CAJ15
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT

CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK4
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT

LCK5
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK6
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK7
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK8
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK9
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC

CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
LCK10
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
GTTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCACAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CATC1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CATC2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
CATC3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT

ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAATAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
EFSGR1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
EFSGR3
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCAGCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAATTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
NFSGR2
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA
CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT
WFSGR1
CATCGGTCTTGTAATCCGAAGATCGAGGGTTAAACCCCCTCCTAGCGCCCAGAAAAGAGAGATTCTAACTCTCACCC
CTGGCTCCCAAAGCCAGAATTCTAAATTAAACTATCTTCTGATAGGAACCTATATGGTCCGGTACCACGTGTAGCATC
ACATGCGTACAGTACACATATCTGGTTCAACACATACATAATATTATTCGTAATATTGTGTGTGTTGTGTTAGTGCATAT
ATATGTATTATCACCATTCATTTATCTTAACCTAAAAGCAAGTACTAACATCCAAGACGTACATAGACCAAATCGTTAAA
ACTCACAAATAAATTATTATAACCTGGGAACTATATTATTCCCCTAGATATGGCACACAATACTTTCCTTGAAATAAACA

CCTAAGGTTTATCATAACCATATTAATGTAGTAAGAGACCACCAACCGGTTCATATAAGGCATTCTATTCATGATAGAA
TCAGGGACACAATATGTGGGGGTAGCACACTGTGAATTATTCCTTGCATCTGGTTCCTATTTCAGGTGTATATTTATAA
GACTCCACCCTCGGATAATTATACTGGCATCTGATTAATGGTGTAATTACATACTCCTCGTTACCCAACATGCCGGGC
ATTCTTTTATATGCATAGGGGTTCTCTTTTTTGGTTTCCTTTCACTTTACATCTCAGAGTGCAGGCGCAATTAACATCTC
AAGGTGGTACTTTTCCTTGCACGAATTAAAGTAGGTTCATCATTAAAAGACATAACTTAAGAATTACATTTTACTCTATC
AAGTGCATAACATATACATCCTTTCTTCAACTTACCCTGATATATACGCCCCCTCTTTTGGTTTCCACGCGTCAAACCC
CCCTACCCCCCTACGCTCAGCAAATCCTGTTCTCCTTGTCAAACCCCGAAACCAAGGAAGGCTCGAGAACGTACGAAC
TAACAAGTTGAGATATCCGTTAGCCATCCGCGTTATATATATATACATACATGTCATATCGACCCCCCCTAAAAATTCT
CTCAAAAATAGCCCAAAAAATTCTACTAAATTTACTGGTAAATTTCTCAATGCTAAAAATTCGAACGTATTTAACCGCTA
GCGTAGCTTAATGCAAAGCGTAACACTGAAGATGT

;
End;

REFERENCES

Alves, M.J., H. Coelho, M.J. Collares-Pereira, and M.M. Coelho. 2001. Mitochondrial DNA variation in the highly endangered cyprinid fish *Anaecypris hispanica*: importance for conservation. Heredity 87: 463 – 473.

Aquadro, D.F., and B.D. Greenberg. 1983. Human mitochondrial DNA variation and evolution: an analysis of nucleotide sequences from seven individuals. Genetics 103: 287 – 312.

Avise, J.C. 2000. Phylogeography: The history and formation of species. Harvard University Press. Cambridge, MA.

Avise, J.C., J. Arnold, R.M. Ball, E. Bermingham, T. Lamb, J.E. Neigel, C.A. Reeb, and N.C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics 18: 489 – 522.

Avise, John C. 1998. The history and purview of phylogeography: a personal reflection. Molecular Ecology 7: 371 – 379.

Baker, A.J., and H.D. Marshall. 1997. Mitochondrial control region sequences as tools for understanding evolution. In: Avian Molecular Evolution and Systematics, D.P. Mindell (ed.). New York: Academic Press, pp. 51 – 82.

Ballard, J., and M. Kreitman. 1995. Is mitochondrial DNA a strictly neutral marker?, Trends Eco. Evol. 10(12): 485 - 488,

Barrett, P. 2003. *Rhinichthys osculus* photograph. USFWS.

Bernatchez, L., and C.C. Wilson. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. Mol. Ecol. 7: 431 – 452.

Birky, C.W., Jr., T. Maruyama, and P. Fuerst. 1983. An approach to population and evolutionary theory for genes in mitochondria and chloroplasts, and some results. Genetics 103: 513 – 527.

Bogengagen, D. and D.A. Clayton. 1974. The number of mitochondrial DNA genomes in mouse L- and human HeLa cells: quantitative isolation of mitochondrial DNA. Journal of Biological Chemistry 249: 7991 – 7995.

Broughton, E.B., and T.E. Dowling. 1997. Evolutionary dynamics of tandem repeats in the mitochondrial DNA control region of the minnow Cyprinella spiloptera. Mol. Biol. Evol. 14(12):1187 – 1196.

Brown, G.G, G. Gadaleta, G. Pepe, C. Saccone, and E. Sbisa. 1986. Structural conservation and variation in the d-loop containing region of vertebrate mitochondrial DNA. J. Mol. Biol. 192: 503 – 511.

Brown, J.R., K. Beckenbach, A.T. Beckenbach, and M.J. Smith. 1996. Length variation, heteroplasmy and sequence divergence in the mitochondrial DNA of four species of sturgeon (*Acipenser*). Genetics 142: 525 – 535.

Bunn, S.E., and J.M. Hughes. 1997. Dispersal and recruitment in streams: evidence from genetic studies. Journal of the North American Benthological Society 16: 338 – 346.

Burton, R.S., and B.N. Lee. 1994. Nuclear and mitochondrial gene genealogies and allozyme polymorphism across a major phylogeographic break in the

marine copepod *Tigriopus californicus*. Proceedings of the National Academy of Sciences 91: 5197 – 5201.

Calsbeek, R., J.N. Thompson, and J.E. Richardson. 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: the California floristic province. Molecular Ecology 12: 1021 – 1029.

Cantatore, P., M. N. Gadaleta, M. Roberti, C. Saccone, and A. C. Wilson. 1987. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature 329: 853 – 855.

Carson, H.L., and A.R. Templeton. 1984. Genetic revolutions in relation to speciation phenomena: The founding of new populations. Annual Review of Ecology and Systematics 15: 97 – 131.

Chatzimanolis, S. and M.S. Caterino. 2007. Toward a better understanding of the "Transverse Range Break;" lineage diversification in southern California. Evolution 61 (9): 2127 – 2147.

Cheng, Y., T. Xu, G. Shi, and R. Wang. 2010. Complete mitochondrial genome of the miiuy croaker *Miichthys miiuy* (Perciformes, Sciaenidae) with phylogenetic consideration. 2010. Marine Genomics 3: 201 – 209.

Chenoweth, S.F., J.M. Hughes, C.P. Keenan, and S. Lavery. 1998. Concordance between dispersal and mitochondrial gene flow: isolation by distance in a tropical teleost, *Lates calcarifer* (Australian barramundi). Heredity 80: 187 – 197.

Clayton, D.A. 1984. Transcription of mammalian mitochondrial genome. Ann. Rev. Biochem. 53: 573 – 594.

Clayton, D.A. 2000. Vertebrate mitochondrial DNA – a circle of surprises. Experimental Cell Research 255: 4 – 9.

Clement, M., D. Posada, and K.A. Crandall. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9: 1657 – 1659.

Conner, J.K., and D.L. Hartl. 2004. A Primer of Ecological Genetics. Sinauer Associates, Inc. Sunderland, MA.

Cooke, G.M., N.L. Chao, and L.B. Beheregaray. 2012. Divergent natural selection with gene flow along major environmental gradients in Amazonia: insights from genome scans, population genetics and phylogeography of the characin fish *Triportheus albus*. 2012. Molec. Ecol. 21: 2410 – 2427.

Cornelius, R.H. 1969. The systematics and zoogeography of *Rhinichthys osculus* (Girard) in Southern California. Unpublished masters thesis, California State University, Fullerton.

Costedoat, C., R. Chappaz, B. Barascud, O. Guillard, and A. Gilles. 2006. Heterogeneous colonization pattern of European Cyprinids, as highlighted by the dace complex (Teleostei: Cyprinidae). Mol. Phylog. Evol. 41: 127 – 148.

Culver, G.B., and C.L. Hubbs. 1917. The fishes of the Santa Ana system of streams in southern California. Lorquinia 1 (11): 82 – 83.

Darwin, C. 1859. On the origin of species by means of natural selection. John Murray, London.

Dawid, I.B., and A.W. Blackler. 1972. Maternal and cytoplasmic inheritance of mitochondrial DNA in *Xenopus*. Dev. Biol. 29: 152 – 161.

Dawson, M.N., J.L. Staton, and D.K. Jacobs. 2001. Phylogeography of the tidewater goby, *Eucyclogobius Newberryi* (Teleostei, Gobiidae), in coastal California. Evolution 55(6): 1167 – 1179.

Delarbre, C., N. Spruyt, C. Delmarre, C. Gallut, V. Barriel, P. Janvier, V. Laudet, and G. Gauchelin. 1998. The complete nucleotide sequence of the mitochondrial DNA of the dogfish, *Scyliorhinus canicula*. Genetics 150: 331 – 344.

DeSalle, R., T. Freedman, E.M. Prager, and A.C. Wilson. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. 26: 157 – 164.

Doda, J.N., C.T. Wright, and D.A. Clayton. 1981. Elongation of displacement loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. USA 78: 6116 – 6120.

Excoffier, L., P.E. Smouse, and J.M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131: 479 – 491.

Faber, J.E., and C.A. Stepien. 1998. Tandemly repeated sequences in the mitochondrial DNA control region and phylogeography of the pike-perches Stizostedion. Mol. Phyl. Evol. 10(3): 310 – 322.

Feeney, R.F., and C.C. Swift. 2008. Description and ecology of larvae and juveniles of three native cypriniforms of coastal southern California. Ichthyol. Res. 55: 65 – 77.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Bio. 27(4): 401 - 410.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39: 783 – 791.

Felsenstein, J. Inferring phylogenies. 2004. Sinauer Associates. Sunderland, MA.

Fisher, R. 1930. The genetical theory of natural selection. The Clarendon Press. United Kingdom.

Flot, J.F., and S. Tillier. 2007. The mitochondrial genome of *Pocillopora* (Cnidaria: Scleractinia) contains two variable regions: The putative D-loop and a novel ORF of unknown function. Gene 401: 80 – 87.

Forister, M.L., J.A. Fordyce, and A.M. Shapiro. 2004. Geological barriers and restricted gene flow in the holarctic skipper *Hesperia comma* (Hesperiidae). Molecular Ecology 13: 3489 – 3499.

Fraser, B.A., N.E. Mandrak, and R.L. McLaughlin. 2005. Lack of morphological differentiation in eastern (*Rhinichthys atratulus*) and western (*Rhinichthys obtusus*) blacknose dace in Canada. Can. J. Zool. 83: 1502 – 1509.

Fresco, J.R., and B.M. Alberts. 1960. The accommodation of noncomplementary bases in helical polyribonucleotides and deoxyribonucleic acids. Proc. Nat. Acad. Sci. USA 46: 311 – 321.

Froufe, E., I. Knizhin, and S. Weiss. 2005. Phylogenetic analysis of the genus *Thymallus* (grayling) based on mtDNA control region and ATPase 6 genes, with inferences on control region constraints and broad-scale Eurasian phylogeography. Mol. Phylog. Evol. 34: 106 – 117.

Fry, A. 1999. Mildly deleterious mutations in avian mitochondrial DNA: evidence from neutrality tests. Evolution 53(5): 1617 - 1620.

Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147: 915 – 925.

Futuyma, D.J. 2005. Evolution. Sinauer Associates, Sunderland, MA.

Guo, X., S. Liu, and Y. Liu. 2003. Comparative analysis of the mitochondrial DNA control region in cyprinids with different ploidy level. Aquaculture 224: 25 – 38.

Haney, R. A., B. R. Silliman, and D. M. Rand. 2010. Effects of selection and mutation on mitochondrial variation and inferences of historical population expansion I a Caribbean reef fish. Molecular Phylogenetics and Evolution 57: 821 – 828.

Harden, D.R. California Geology. 2004. Pearson Education, Inc. Upper Saddle River, NJ.

Harpending, H.C., M. A. Batzer, M. Gurven, L.B. Jorde, and A.R. Rogers. 1998. Genetic traces of ancient demography. Proc. Natl. Acad. Sci. U.S.A. 95: 1961 – 1967.

Harrison, R.G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends in Ecology and Evolution 4: 6 – 11.

Hayashi, J.I., Y. Tagashira, and M.C. Yoshida. 1985. Absence of extensive recombination between inter- and intraspecies mitochondrial DNA in mammalian cells. Exp. Cell. Res. 160: 387 – 395.

Hennig, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana, Illinois.

Hickman, J.C.(editor). The Jepson manual: higher plants of California. 1993. U.C. Press. Berkeley, CA.

Hocutt C.E. (editor). The zoogeography of North American freshwater fishes. 1986. John Wiley & Sons. New York, NY.

Hoeh, W.R., K.H. Blakley, and W.M. Brown. 1991. Heterplasmy suggests limited biparental inheritance of *Mytilis* mitochondrial DNA. Science 251: 1488 – 1490.

Hoelzel, A.R., J.M. Hancock, and G.A. Dover. 1991. Evolution of the cetacean mitochondrial d-loop region. Mol. Biol. Evol. 8(3): 475 – 493.

Hoelzel, A.R., J.V. Lopez, G.A. Dover, and S.J. O'Brien. 1994. Rapid evolution of a heteroplasmic repetitive sequence in the mitochondrial DNA control region of carnivores. J. Mol. Evol. 39: 191 – 199.

Hubbs, C.L., and R.R. Miller. 1948. Correlation between fish distribution and the hydrographic history in the desert basins of western United States. In: The Great Basin, with emphasis on Glacial and post-glacial times. Pp 17 – 144. Bull. Univ. Utah 38, Biol. Ser. 10(7).

Hubbs, C.L., Miller, R.R., and Hubbs. L.C. 1974. Hydrographic history and relict fishes of the north-central great basin. Memoirs of the California Academy of Sciences, Vol. 7. San Francisco, CA.

Hudson, R.R., M. Slatkin, and W.P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132: 583 – 589.

Huelsenbeck, J.P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754 – 755.

Hughes, J.M., S.E. Bunn, C. Cleary, and D.A. Hurwood. 2000. A hierarchical analysis of the genetic structure of an aquatic insect *Bungona* (Baetidae: Ephemeroptera). Heredity 85: 561 – 570.

Hurst, D.H., S.E. Bartlett, W.S. Davidson, and I.J. Bruce. 1999. The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. Gene 239: 237 – 242.

Hurwood. D.A., and J.M. Hughes. 2001. Nested clade analysis of the freshwater shrimp, *Caridina zebra* (Decapoda: Atyidae) from north-eastern Australia. Mol. Ecol. 10: 113 – 125.

Jemt, E., O. Persson, Y. Shi, M. Mehmedovic, J.P. Uhler, M.D. Lopez, C. Freyer, C.M. Gustafsson, T. Samuelsson, and M. Falkenberg. 2015. Regulation of DNA replication at the end of the mitochondrial d-loop involves the helicase TWINKLE and a conserved sequence element. Nuc. Acids Res. 43(19): 9262 – 9275.

Jordan, D.S., and B.W. Evermann. 1896. The fishes of North and Middle America: a descriptive catalogue of the species of fish-like vertebrates found in the waters of North America, north of the Isthmus of Panama. Bull. U.S. Natl. Mus. 47: 1 – 1240.

Jukes, T.H., and C.R. Cantor. 1969. Evolution of protein molecules. New York: Academic Press. pp. 21–132.

Keeney, D.B., M.R. Heupel, R.E. Hueter, and E.J. Heist. 2005. Microsatellite and mitochonndrial DNA analyses of the genetic structure of blacktip shark (*Carcharhinus limbatus*) nurseries in the northwestern Atlantic, Gulf of Mexico, and Caribbean Sea. Molec. Ecol. 14: 1911 – 1923.

Kimura, M. 1968. Evolutionary rate at the molecular level. Nature 217: 624 – 626.

Kuchta, S.R., and A. Tan. 2006. Lineage diversification on an evolving landscape: phylogeography of the California newt, *Taricha torosa* (Caudata: Salamandridae). Biol. J. of Linn. Soc. 89: 213 – 239.

Kvist, L., M. Ruokonen, A. Thessing, J. Lumme, and M. Orell. 1998. Mitochondrial control region polymorphism reveal high amount of gene flow in Fennoscandian willow tit (*Parus montanus* borealis). Hereditas 128: 133 – 143.

LaRivers, I. 1952. A key to Nevada Fishes. S. Calif. Acad. Sci. Bull. 51(3): 86 – 102.

Lee, D.S., C.R. Gilbert, C.H. Hocutt, R.E. Jenkins, D.E. McCallister, J.R. Stauffer Jr. 1980. Atlas of North American freshwater fishes. North Carolina State Museum of Nat. History, Raleigh NC.

Lee, J., M. Miya, Y. Lee, C.G. Kim, E. Park, Y. Aoki, and M. Nishida. 2001. The complete DNA sequence of the mitochondrial genome of the self-fertilizing fish *Rivulus marmoratus* (Cyprinodontiformes, Rivulidae) and the first description of duplication of a control region in fish. 2001. Gene 280: 1 – 7.

Lee, W., J. Conroy, W.H. Howell, and T.D. Kocher. 1995. Structure and evolution of Teleost mitochondrial control regions. J. Mol. Evol. 41: 54 – 66.

Lemmon, E.M., A.R. Lemmon, and D.C. Cannatella. 2007. Geological and climatic forces driving speciation in the continentally distributed trilling chorus frogs (*Pseudacris*). Evol. 61(9): 2086 – 2103.

Li, H.C., J.L. Bischoff, T.L. Ku, and Z.Y. Zhu. 2004. Climate and hydrology of the last interglaciation (MIS 5) in Owens Basin, California: isotopic and geochemical evidence from core OL-92. Quaternary Sci. Rev. 23: 49 – 63.

Lomolino, M.V., B.R. Riddle, and J.H. Brown. 2006. Biogeography, 3rd ed. Sinauer Associates. Sunderland, MA.

Lunt, D.H., and B.C. Hyman. 1997. Animal mitochondrial DNA recombination. Nature 387: 247.

Lynch, M., and T.J. Crease. 1990. The analysis of population survey data on DNA sequence variation. Molecular Biology and Evolution 7: 377 – 394.

Macey, J.R., J.L. Strasburg, and J.A. Brisson. 2001. Molecular phylogenetics of western North American frogs of the Rana boylii species group. Mol. Phyl. Evol. 19: 131 – 143.

Machordom. A., and I. Doadrio. 2001. Evidence of a Cenozoic Betic-Kabilian connection based on freshwater fish phylogeography (*Luciobarbus,* Cyprinidae). Mol. Phyl. Evol. 18(2): 252 – 263.

Magoulas, A., and E. Zouros. 1993. Restriction-site heteroplasmy in anchovy (*Engraulis encrasicolus*) indicates incidental biparental unheritance of mitochondrial DNA. Mol. Biol. Evol. 10: 319 – 325.

Maldonado, J.E., C. Vila, and R.K. Wayne. 2001. Tripartite genetic subdivisions in the ornate shrew (*Sorex ornatus*). Mol. Ecol. 10: 127 – 147.

Mayden, R.L., and E.O. Wiley. 1992. Phylogenetic systematics. In: Systematics, historical ecology, and North American freshwater fish (Ed: R.L. Mayden). Stanford University Press, Stanford, California.

Mayr, E. 1942. Systematics and Origin of Species. N.Y. Columbia University Press.

McConnell, S.D.J. 2004. Mapping aquatic faunal exchanges across the Sunda shelf, south-east Asia, using distributional and genetic data sets from the cyprinid fish *Barbodes gonionotus*. J. Nat. His. 38: 651 – 670.

McGinnis, S.M. Field guide to freshwater fishes of Southern California. 2006. University of California Press. Berkeley, CA.

McMillan, W.O., and S.R. Palumbi. 1997. Rapid rate of control-region evolution in pacific butterflyfishes (Chaetodontidae). J. Mol. Evol. 45: 473 – 484.

Meffe, G.K., and R.C. Vrijenhoek. 1988. Conservation genetics in the management of desert fish. Conservation Biology 2: 157 – 168.

Metcalf, A.E., L. Nunney, and B.C. Hyman. 2001. Geographic patterns of genetic differentiation within the restricted range of the endangered Stephens' kangaroo rat *Dipodomys stephensi*. Evolution 55: 1233 – 1244.

Miller, R.R. 1946. On relations between fish distribution and Pleistocene hydrography in eastern California and southwestern Nevada, with a map of the Pleistocene waters. J. Geology. LIV 1: 43 – 53.

Minckley, W.L, Hendrickson, D.A., and C.E. Bond. 1986. Geography of western North American freshwater fishes: description and relationships to intracontinental tectonism. In Hocutt C.H., Wiley, E.O. (editors). The zoogeography of North American freshwater fishes. 1986. John Wiley & Sons. New York, NY.

Moore, W.S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial - gene trees versus nuclear – gene trees. Evolution 49: 718 – 726.

Moritz, C. 1995. Uses of Molecular Phylogenies for Conservation. Phil. Trans. R. Soc. Lond. B. 349: 113-118.

Moritz, C., T.E. Dowling, and W.M. Brown. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ecological Systematics 18: 269 – 292.

Moyle, P. 1995. Conservation of native freshwater fishes in the mediterranean-type climate of California, USA: A review. Bio. Cons. 72(2): 271 - 279.

Moyle, P.B. 2002. Inland fishes of California. University of California Press. Berkeley, CA

Myers, N., R.A. Mittermeier, C.G. Mittermeier, G.A.B. deFonseca, and J. Kent. 1999. Biodiversity hotspots for conservation priorities. Nature 403: 853 – 858.

Nachman, N.W., and S.L. Crowell. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics. 156 (1): 297–304.

Nerkowski, S. 2015. Microsatellite analysis of population structure in the Santa Ana speckled dace (Rhinichthys osculus). Unpublished master's thesis. California State University, San Bernardino.

Oakey, D.D., M.E. Douglas, and M.R. Douglas. 2004. Small fish in a large landscape: diversification of *Rhinichthys osculus* (Cyprinidae) in western North America. Copeia 2: 207 – 221.

Palumbi, S.R., and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequence sequences and mtDNA of humpback whales. Molecular Biology and Evolution 11: 426 – 435.

Peakall, R., and P.E. Smouse. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6: 288 – 295.

Peakall, R., and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. Bioinformatics 28: 2537 – 2539.

Perdices, A., D. Sayanda, and M.M. Coelho. 2005. Mitochondrial diversity of *Opsariichthys bidens* (Teleostei, Cyprinidae) in three Chinese drainages. Mol. Phyl. Evol. 37: 920 – 927.

Pesole, G., Gissi, C., Chirico, D.E., Saccone, C. 1999. Nucleotide substitution rate of mammalian mitochondrial genomes. J. Mol. Evol. 48: 427 – 434.

Pfrender, M.E., J. Hicks, and M. Lynch. 2004. Biogeographic patterns and current distribution of molecular-genetic variation among populations of speckled dace, *Rhinichthys osculus*. Molecular Phylogenetics and Evolution 30: 490 – 502.

Phillipsen, I.C. 2005. Phylogeography of a stream-dwelling frog (*Pseudacris cadaverina*) in southern California. Unpublished masters thesis, California State University, San Bernardino.

Phillipsen, I.C., and A.E. Metcalf. 2009. Phylogeography of a stream-dwelling frog (*Pseudacris cadaverina*) in southern California. Molecular Phylogenetics and Evolution 53(1): 152 – 170.

Posada, D., and K.A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817 – 818.

Preziosi, R.F., and D.J. Fairbairn. 1992. Genetic population structure and levels of gene flow in the stream dwelling waterstrider *Aquarius* (= Gerris) *remgis* (Hemiptera: Gerridae). Evolution 46.

Quinn, T.W. 1992. The genetic legacy of mother goose – phylogeographic patterns of lesser snow goose *Chen caeruluscens caerulescens* maternal lineages. Mol. Ecol. 1: 105 – 117.

Randi, E., and Lucchini, V. 1998. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. J. Mol. Evol. 47: 149 – 157.

Ray, A.R., and L. Densmore. 2002. The crocodilian mitochondrial control region: general structure, conserved sequences, and evolutionary implications. J. Exp. Zoo. (Mol. Dev. Evol.) 294: 334 – 345.

Recuero, E., I. Martinez-Solano, G. Parra-Olea, and M. Garcia-Parra. 2006. Phylogeography of *Pseudacris regilla* (Anura: Hylidae) in western North America, with a proposal for a new taxonomic rearrangement. Mol. Phyl. Evol. 39: 293 – 304.

Robalo, J.I., C.S. Santos, V.C. Almada, and I. Doadrio. 2006. Paleobiogeography of two Iberian endemic cyprinid fishes (*Chondrostoma arcasii-Chondrostoma macrolepidotus*) inferred from mitochondrial DNA sequence data. J. Heredity 97(2): 143 – 149.

Rodriguez-Robles, J.A., D.F. Denardo, and R.E. Staub. 1999. Phylogeography of the California mountain kingsnake, *Lampropeltis zonata* (Colubridae). Mol. Ecol. 8: 1923 – 1934.

Rodriguez-Robles, J.A., G.R. Stewart, and T.J. Papenfuss. 2001. Mitochondrial DNA-based phylogeography of North American rubber boas, *Charina bottae* (Serpentes: Boidae). Mol. Phylogenet. Evol. 18: 227 – 237.

Rokas, A., E. Ladoukakis, and E. Zouros. 2003. Animal mitochondrial DNA recombination revisited. Trends in Ecology and Evolution. 18: 411 – 417.

Rogers, A.R., and H. Harpending. 1992. Population growth makes waves in the distribution of pair wise genetic differences. Mol. Biol. Evol. 9: 552 – 569.

Ronquist, F., M. Teslenko, P. Van Der Mark, D.L. Ayres, A. Darling, S. Hhna, B. Larget, L. Liu, M.A. Suchard , and J. Huelsenbeck. 2012. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61: 539 – 542.

Rosen, D.E. 1978. Vicariant patterns and historical explanation in biogeography. Syst. Zool. 27: 159 – 188.

Ruokonen, M., and L. Kvist. 2002. Structure and evolution of the avian mitochondrial control region. Mol. Phylog. Evol. 23: 422 – 432.

Ryder, O.A. 1986. Species conservation and systematics: the dilemma of subspecies. Trends Ecol. Evol. 1: 9 – 10.

Saccone, C., M. Attimonelli, and E. Sbisa. 1987. Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. J. Mol. Evol. 26: 205 – 211.

Saccone, C., G. Pesole, and E. Sbisa. 1991. The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. J. Mol. Evol. 33: 83 – 91.

Sanger, F., S. Nicklen, and A.R. Coulsen. 1977. DNA Sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463 – 5467.

Santa Ana Speckled Dace Recovery Project. 2005. United States Forest Service and California Department of Fish and Game Bi-Annual Report. Kerwin Russell, editor. April – October 2005.

Santa Ana River Watershed Project Authority (SAWPA). 2004. Old, Grand Prix, and Padua Fires (October, 2003): Burn Impacts to Water Systems and Resources. Santa Ana River Watershed Area. San Bernardino National Forest, California. Santa Ana Watershed Project Authority. Summary Report. Riverside, California.

Sbisa, E., F. Tanzariello, A. Reyes, G. Pesole, and C. Saccone. 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene 205: 125 – 140.

Scharf, S.J., G.T. Horn, and H.A. Erlich. 1986. Direct cloning and sequence analysis of enzymatically amplified genomic sequences. Science 233: 1076 – 1078.

Scheffler, I.E. 1999. Mitochondria. Wiley-Liss. New York, NY.

Schneider, S. and L. Excoffier. 1999. Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. Genetics. 152 (3): 1079 – 89.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: a software for population genetics data analysis. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva. Geneva, Switzerland.

Schneider-Broussard, R., and J.E. Neigel. 1997. A large-subunit mitochondrial DNA sequence translocated to the nuclear genome of two stone crabs (*Menippe*). Mol. Biol. Evol. 14: 156 – 165.

Seeliger, L. 1945. Variation in the Pacific mud turtle. Copeia, 1945(3): 150 - 159.

Sgariglia, E.A. and K.J. Burns. 2003. Phylogeography of the California thrasher (*Toxostoma redivivum*) based on nested-clade analysis of mitochondrial-DNA variation. Auk 120: 346 - 361.

Smith, G.R., and T.E. Dowling. 2008. Correlating hydrographic events and divergence times of speckled dace (Rhinichthys: *Teleostei: Cyprinidae*) in the Colorado River drainage. 2008. In: Late Cenozoic drainage history of the southwestern Great Basin and lower Colorado River region: geologic

and biotic perspectives. (Reheis, M.C, R. Hershler, and D.M. Miller, eds). The Geological Society of America, special paper 439. Boulder, Colo. pp 301 - 317.

Southern, S.O., P.J. Southern, and A.E. Dizon. 1988. Molecular characterization of a cloned dolphin mitochondrial genome. J. Mol. Evol. 28: 32 – 40.

Spencer, J.E., G.R. Smith, and T.E. Dowling. 2008. Middle to late Cenozoic geology, hydrography, and fish evolution in the American southwest. In: Late Cenozoic drainage history of the southwestern Great Basin and lower Colorado River region: geologic and biotic perspectives. (Reheis, M.C, R. Hershler, and D.M. Miller, eds). The Geological Society of America, special paper 439. Boulder, Colo. pp 279 – 299.

Spinks, P.Q. and H.B. Shaffer. 2005. Range-wide molecular analysis of the western pond turtle (*Emys marmorata*): cryptic variation, isolation by distance, and their conservation implications. Molecular Ecology 14: 2047 – 2064.

Stefani, F., P. Galli, S. Zaccara, and G. Crosa. 2004. Genetic variability and phylogeography of the cyprinid *Telestes muticellus* within the Italian peninsula as revealed by mitochondrial DNA. J. Zoo. Syst. Evol. Res. 42: 323 – 331.

Stepien, C.A., and J.E. Faber. 1998. Population genetic structure, phylogeography and spawning philopatry in walleye (*Stizostedion vitreum*)

from mitochondrial DNA control region sequences. Mol. Ecol. 7: 1757 – 1769.

Stepien, C.A., and T.D. Kocher. 1997. Molecules and morphology in studies of fish evolution. In “Molecular Systematics of Fishes” (T.D. Kocher and C.A. Stepien, Eds.). Pp 1-12, Academic Press, San Diego.

Stephenson, J.R., and G.M. Calcarone. 1999. Southern California mountains and foothills assessment: habitat and species conservation issues. Pacific Southwest Research Station. Albany, CA.

Stewart, D.T., and A.J. Baker. 1994. Patterns of sequence variation in the mitochondrial D-loop region of shrews. Mol. Biol. Evol. 11: 9 – 21.

Swofford, D.L., G.J.Olson, P.J. Waddell, and D.M. Hillis. 1996. Phylogenetic Inference. Pp. 407 – 514 in Molecular Systematics (D.M. Hillis, C. Moritz, and B.K. Mable, eds.). Sinauer Assoc., Sunderland, MA.

Swofford, D.L. 2003. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Inc. Sunderland, MA.

Tadao, K., Y. Motoi, K. Masaaki, and O. Toshio. 2001. Population structure and local differentiation in the delicate loach, *Niwaella delicate*, as revealed by mitochondrial DNA and morphological analysis. Ichthyol. Res. 48: 127 – 135.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585 – 595.

Takahashi, H., and A. Goto. 2001. Evolution of East Asian Ninespine Sticklebacks as shown by mitochondrial DNA control region sequences. Mol. Phyl. Evol. 21(1): 135 - 155.

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Molecular Biology and Evolution. 9 (4): 678 – 687.

Tamura, K., and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512 – 526.

Tamura, K., G. Stecher, D. Peterson, A. Filipski, and S. Kumar. 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Bio. Evol. 30: 2725 – 2729.

Tan, A.M., and D.B. Wake. 1995. MtDNA phylogeography of the California newt, *Taricha torosa* (Caudata, Salamandridae). Molecular Phylogenetics and Evolution 4: 383 – 394.

Tang, Q., H. Liu, R. Mayden, and B. Xiong. 2006. Comparison of evolutionary rates in the mitochondrial DNA cytochrome *b* gene and control region and their implications for phylogeny of the Cobitoidea (Teleostei: Cypriniformes). Mol. Phylogen. Evol. 39: 347 – 357.

Templeton, A.R. 1980. The theory of speciation via the founder principle. Genetics 94: 1011 – 1038.

Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin, and D.G. Higgins. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876 – 4882.

Thompson, R.S., K. Whitlock, P.J. Bartlein, S.P. Harrison, and W.G. Spaulding. 1993. Climatic changes in the western United States since 18,000 yr. b.p. In H.E. Wright, J.E. Kutzbach, T. Webb, W.F. Ruddiman, F.A. Street-Perrott and P.J. Bartlein, eds. Global climates since the last glacial maximum. University of Minnesota Press. Minneapolis, MN.

Walberg, M.W., and D.A. Clayton. 1981. Sequence and properties of the human KB cell and mouse L cell d-loop regions of mitochondrial DNA. Nucleic Acids Res. 9: 5411 – 5421.

Walker, D., P.E. Moler, K.A. Buhlmann, and J.C. Avise. 1998. Phylogeographic uniformity in mitochondrial DNA of the snapping turtle (*Chelydra serpentina*). Anim. Conserv. 1: 55 – 60.

Waples, R. S. 1995. Evolutionary significant units and the conservation of biological diversity under the Endangered Species Act. In: Evolution and the Aquatic Ecosystem: Defining Unique Units in Population Conservation. American Fisheries Society Symposium 17. Bethesda, Maryland.

Weiss, S., T. Kopun, and S. Bajec. 2013. Assessing natural and disturbed population structure in European grayling *Thymallus thymallus*: melding

phylogeographic, population genetic and jurisdictional perspectives for conservation planning. Journal of Fish Biology 82: 505 – 521.

White, T.J., and N. Arnheim, and H.A. Erlich. 1989. The polymerase chain reaction. Trends Genet. 5: 185 – 188.

Wilson, A.C., R.L. Cann, S.M. Carr, M. George, U. B. Gyllensten, K. M. Helm-Bychowski, R. G. Higuchi, S. R. Palumbi, E. M. Prager, R. D. Sage, and M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. Biol. J. Linn. Soc. 26: 375 – 400.

Wolstenholme, D.R., and D.O. Clary. 1985. Sequence evolution of *Drosophila* mitochondrial DNA. Genetics 109: 725 – 744.

Woodman, D.A. 1992. Systematic relationships within the cyprinid genus *Rhinichthys*. In: Mayden, R.L. (Ed.), Systematics, Historical Ecology, and North American Freshwater Fishes. Stanford University Press, Stanford, CA. pp. 374 – 391.

Wright, S. 1931. Evolution in Mendelian populations. Genetics. 16: 97 – 159.

Yokoyama, R. and A. Goto. 2002. Phylogeography of a freshwater sculpin, *Cottus nozawae*, from the northeastern part of Honshu Island, Japan. Ichthyol. Res. 49: 147 – 155.

Zardoya, R., and I. Doadrio. 1998. Phylogenetic relationships of Iberian cyprinids: systematic and biogeographical implications. Proc. R. Soc. Lond. B 265: 1365 – 1372.

Zhang, D., and G.M. Hewitt. 1997. Insect mitochondrial control region: a review of its structure, evolution, and usefulness in evolutionary studies. Biochem. Sys. And Ecol. 25(2): 99 – 120.

Zhang, D., J.M. Szymura, and G.M. Hewitt. 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. J. Mol. Evol. 40: 382 – 391.

Zhu, D., B.G.M. Jamieson, A. Hugall, and C. Moritz. 1994. Sequence evolution and phylogenetic signal in control-region and cytochrome b sequences of rainbow fishes (*Melanotaeniidai*). Mol. Biol. Evol. 11(4): 672 – 683.

Zouros, E., K.R. Freeman, A. Oberhauser-Ball, and G.H. Pogson. 1992. Direct evidence for extensive paternal mitochondrial DNA inheritance in the mussel *Mytilus*. Nature 359: 412 – 414.